

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 20:33:53; Search time 1379.4 Seconds
(without alignments)
223.078 Million cell updates/sec

Title: US-09-803-736-1079_COPY_13662_13680

Perfect score: 19

Sequence: 1 acaagatagacagacacatgc 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST.*
1: em_estba.*
2: em_estlum.*
3: em_estin.*
4: em_estma.*
5: em_estov.*
6: em_estpl.*
7: em_estro.*
8: em_estro.*
9: gb_estl.*
10: gb_estl2.*
11: gb_hic.*
12: gb_est3.*
13: gb_est4.*
14: gb_est5.*
15: em_estlum.*
16: em_estom.*
17: gb_gss.*
18: em_gss_hum.*
19: em_gss_inv.*
20: em_gss_pln.*
21: em_gss_frt.*
22: em_gss_fun.*
23: em_gss_mam.*
24: em_gss_mus.*
25: em_gss_other.*
26: em_gss_pro.*
27: em_gss_rod.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	17.4	91.6	194	9	AI621482 486090H06
2	17.4	91.6	243	10	BA423297 BA423297
3	17.4	91.6	438	13	BI993076 1020072D0
4	17.4	91.6	530	12	BG518007 947065F11
5	17.4	91.6	672	17	AO324775 mxxb0019D
6	17	89.5	235	9	AI383272 tc77c07.x

Result No.	Score	Query Match	Length DB	ID	Description
7	17	89.5	561	10	BE581326
8	17	89.5	580	12	BG226225
9	17	89.5	601	12	BG227716
10	17	89.5	721	17	BH202699
11	16.4	86.3	149	10	AM749252
12	16.4	86.3	213	9	AA742620
13	16.4	86.3	261	14	T97764
14	16.4	86.3	309	10	BE044749
15	16.4	86.3	309	17	BH269502
16	16.4	86.3	321	12	BE583721
17	16.4	86.3	324	10	AV884915
18	16.4	86.3	338	17	AF248934
19	16.4	86.3	365	10	AV81832
20	16.4	86.3	386	17	AQ087842
21	16.4	86.3	388	13	B1314938
22	16.4	86.3	422	17	A215531
23	16.4	86.3	426	10	AV819692
24	16.4	86.3	435	10	AM285138
25	16.4	86.3	440	14	T99513
26	16.4	86.3	443	17	AQ760565
27	16.4	86.3	449	9	A1977693
28	16.4	86.3	449	9	AU227899
29	16.4	86.3	451	17	AQ773942
30	16.4	86.3	453	17	B80690
31	16.4	86.3	492	17	A2604563
32	16.4	86.3	496	17	AQ407058
33	16.4	86.3	499	17	AO859604
34	16.4	86.3	517	14	BM959693
35	16.4	86.3	521	17	AQ429168
36	16.4	86.3	542	13	BM181772
37	16.4	86.3	551	17	AQ597790
38	16.4	86.3	560	14	BM961073
39	16.4	86.3	579	14	BM962713
40	16.4	86.3	598	10	AV840381
41	16.4	86.3	598	17	BH355113
42	16.4	86.3	599	9	AA855230
43	16.4	86.3	600	12	BG804059
44	16.4	86.3	604	12	BE572470
45	16.4	86.3	606	12	BG557117

ALIGNMENTS

RESULT 1
AI621482/c
LOCUS 486090H06.xl 486 - leaf primordia cDNA library from Hake lab Zea
DEFINITION mays cDNA, mRNA sequence.
ACCESSION AI621482
VERSION AI621482.1 GI:4630608
KEYWORDS EST.
SOURCE Zea mays.
ORGANISM Zea mays.
REFERENCE Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 194)

REFERENCE 1 (bases 1 to 194)
AUTHORS Walbot,V.
TITLE Maize ESTs from various cDNA libraries sequenced at Stanford
JOURNAL Unpublished (1999)
COMMENT Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 723 8221
Email: walbot@stanford.edu
Plate: 486090 row: H column: 06.
Location/Qualifiers
1..194
/organism="Zea mays"

FEATURES
source

FEATURES	Location/Qualifiers
Db	85
ACCESSION	BA423297
VERSION	BA423297.1
KEYWORDS	EST, house mouse, Mus musculus
SOURCE	Mus musculus
ORGANISM	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE	1 (bases 1 to 243)
AUTHORS	Kono H., Aizawa K., Akahira S., Akiyama J., Arakawa T., Carninci P., Endo T., Fukuda S., Fukunishi Y., Hata A., Hayatsu N., Hirozane T., Hori F., Ishii Y., Ishikawa J., Ishikawa T., Itoh M., Iizawa M., Kadota K., Kagawa I., Kai C., Kawai U., Kikuchi N., Kiyosawa H., Kojima Y., Kondo S., Koya S., Kurihara C., Kusakabe M., Matsuyama T., Miki R., Mizuno Y., Nakamura M., Oda H., Okazaki Y., Ono T., Owa C., Saito H., Sakai C., Sato K., Shibata K., Shibata Y., Shigemoto Y., Shinagawa A., Shiraki T., Sogabe Y., Suganara Y., Suzuki H., Suzuki H., Tagawa A., Takahashi F., Tomianga N., Toya T., Tsunoda Y., Watanishi A., Yokota T., Yoshida K., Yoshiki A., Yoshino M., Muramatsu M. and Hayashizaki Y.
TITLE	Riken Mouse ESTs (Kono H., et al.)
JOURNAL	Unpublished (2000)
COMMENT	Contact: Yoshihide Hayashizaki Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute The Institute of Physical and Chemical Research (RIKEN) 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 220-0045, Japan Tel: 81-45-503-9222 Fax: 81-45-503-9226 Email: genome-resgsc.riken.go.jp, url:http://genome-gsc.riken.go.jp/ Carninci P., Nishiyama Y., Westover A., Itoh M., Nagoka S., Sasaki N., Okazaki Y., Muramatsu M. and Hayashizaki Y. Thermosabilization and thermocycling for the synthesis of full length cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998) Itoh M., Katsunai T., Akiyama J., Shibata K., Iizawa M., Kawai U., Tomaru Y., Carninci P., Shibata Y., Ozawa Y., Muramatsu M., Okazaki Y. and Hayashizaki Y. Automated filtration-based high-throughput plasmid preparation system. Genome Res. 9 (5), 463-470 (1999) Carninci P. and Hayashizaki Y. High-efficiency full-length cDNA cloning. Methods Enzymol. 303, 13-44 (1999) Please visit our web site (http://genome-rtc.riken.go.jp) for further details.

```

source
1. 243
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone_id="C53003607"
/clone_lib="Riken full-length enriched, 12 days embryo
spinal cord"
/tissue_type="spinal cord"
/dev_stage="12 days embryo"
/lab_host="DH10B"
/notes="Site 1: SalI; Site 2: BamHI; cDNA library was
prepared and sequenced in Mouse Genome Encyclopedia
Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in
RIKEN, Division of Experimental Animal Research in Riken
contributed to prepare mouse tissues. 1st strand cDNA was
primed with a primer 15'
GAGGAGAGAGAGGATCCAGAGCGCTTTTGTGTTTTPYVN 3'. cDNA was
prepared by using trehalose thermo-activated reverse
transcriptase and subsequently enriched for full-length by
cap-trapper. Second strand cDNA was prepared with the
primer adapter of sequence 15'
GAGGAGAGAGATCTCGAGTGTATTAATTAATCCCGCCCCC 3'. cDNA
was cleaved with XhoI and BamHI. Vector: a modified
pluescript KS(+) after full excision from Lambda FLC I."

BASE COUNT      64 a      53 g      57 t
ORIGIN

Query Match      91.6%; Score 17.4; DB 10; Length 243;
Best Local Similarity 94.7%; Pred. No. 2.5e+03;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY      1 ACNAGATCAGAACAAATGC 19
||||| |||||||
Db      55 ACNAGATTCAGAACAAATGC 37

RESULT 3
BI993076/c      438 bp      mRNA      linear      EST 24-Oct-2001
DEFINITION      1020072E05.x3 1020 - Unigene II from Maize Genome Project Zea mays
CDNA, mRNA sequence.
ACCESSION      BI993076
VERSION      BI993076.1 GI:16379648
KEYWORDS      EST.
SOURCE      Zea mays.
ORGANISM      Zea mays
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE      1 (bases 1 to 438)
AUTHORS      Walbot V.
TITLE      Maize ESTs from various cDNA libraries sequenced at Stanford
UNIVERSITY
JOURNAL      Unpublished (1999)
COMMENT      Contact: Walbot V
Department of Biological Sciences
Stanford University
855 California Ave, Palo Alto, CA 94304, USA
Tel: 650 723 2227
Fax: 650 725 8221
Email: walbot@stanford.edu
Plate: 1020072 row: E column: 05.
Location/Qualifiers
1. 438
/organism="Zea mays"
/db_xref="taxon:4577"
/clone_id="1020 - Unigene II from Maize Genome Project"
/notes="This library represents the unique genes found in
the second round of EST sequencing at Stanford University
for the maize genome project. Sequences are present from
libraries 947 and 949. Contigs were assembled using
ZMDBS assembler and 2 representatives from each contig were
selected for the Unigene set. All singlets were also

```

FEATURES

Location/Qualifiers

selected for the un

. All singlets were also

BASE COUNT 102 a 138 c 88 g 110 t
 ORIGIN
 Query Match 91.6%; Score 17.4; DB 13; Length 438;
 Best Local Similarity 94.7%; Pred. No. 2.7e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
 |||||
 Db 63 ACAAGATACAGAACAGGC 45

RESULT 4
 B3518007/c 530 bp mRNA linear EST 02-APR-2001
 LOCUS 947065F11.x1 947 - 2 week shoot from Barkan lab Zea mays cDNA, mRNA
 DEFINITION
 sequence.
 ACCESSION B3518007
 VERSION B3518007.1 GI:13515731
 KEYWORDS EST.
 SOURCE Zea mays.
 ORGANISM Zea mays.
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
 clade; Panicoideae; Andropogoneae; Zea.
 1 (bases 1 to 530)
 Malhot V.
 Zea ESTs from various cDNA libraries sequenced at Stanford
 UNIVERSITY
 JOURNAL Unpublished (1999)
 COMMENT Contact: Walbot V
 Department of Biological Sciences
 Stanford University
 855 California Ave, Palo Alto, CA 94304, USA
 Tel: 650 723 8221
 Fax: 650 723 8221
 Email: walbot@stanford.edu
 Plate: 947065 row: F column: 11.
 Location/Qualifiers
 1..530
 /organism="Zea mays"
 /cultivar="B73"
 /db_xref="taxon:4577"
 /clone_lib="947 - 2 week shoot from Barkan lab"
 /tissue_type="leaf and stem, including leaf base"
 /dev_stage="2 week old seedling (3 leaves)"
 /lab_host="X1-Blue"
 /note="Organ: shoot; Vector: Lambda ZAP (pBluescript SK-);
 Site: 1: EcoRI; Site 2: XhoI; Directionally cloned using
 Stratagene's unitap XR cDNA cloning kit with the 5' end
 at the EcoRI site. The library represents 8 x 10e5
 independent recombinant phage. The plants were greenhouse
 grown."
 BASE COUNT 138 a 156 c 104 g 132 t
 ORIGIN
 Query Match 91.6%; Score 17.4; DB 12; Length 530;
 Best Local Similarity 94.7%; Pred. No. 2.7e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
 |||||
 Db 217 ACAAGATACAGAACAGGC 199

RESULT 5
 A0324775 672 bp DNA linear GSS 08-JAN-1999
 LOCUS m9xb0019D12r CUGI Rice Blast BAC Library Magnaporthe grisea genomic
 DEFINITION clone m9xb0019D12r, DNA sequence.
 ACCESSION A0324775
 VERSION A0324775.1 GI:4116627

KEYWORDS GSS.
 SOURCE Magnaporthe grisea.
 ORGANISM Magnaporthe grisea
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
 Sordariomycetes Incertae sedis; Magnaportheaceae; Magnaporthe.
 1 (bases 1 to 672)
 Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,
 Phillips, K., Saslowski, M., Ming, R.A. and Dean, R.A.
 A BAC End Sequencing Framework to Sequence the Magnaporthe grisea
 Genome
 JOURNAL Unpublished (1998)
 COMMENT Contact: Dean RA
 Clemson University Genomics Institute
 Clemson University
 100 Jordan Hall, Clemson University, Clemson, SC 29634
 Tel: 864 656 5737
 Fax: 864 656 4293
 Email: rdean@clemson.edu
 Seq primer: CGAAGACGCTATGACGACG
 Class: BAC ends
 High quality sequence start: 57
 High quality sequence stop: 500.
 Location/Qualifiers
 1..672
 /organism="Magnaporthe grisea"
 /strain="70-15"
 /db_xref="taxon:148305"
 /clone_lib="CUGI Rice Blast BAC library"
 /tissue_type="Protoplasts"
 /lab_host="E. coli DH10B"
 /note="Vector: pBACWICH; Site: 1: HindIII; Site 2: HindIII;
 Rice blast is one of the most devastating fungal diseases
 of rice world wide. It is a filamentous ascomycete with
 a haploid genome (n=7) of approximately 40 Mbp. Rice
 blast is an important model fungal pathogen for studying
 numerous aspects of the fungal-host interaction. In
 order to facilitate genome wide analysis, a BAC library
 containing 9216 clones with an average insert size of 130
 kbp was constructed. This library represents greater
 than 25X genome coverage. High density colony filters
 are available upon request."

BASE COUNT 188 a 148 c 164 g 171 t 1 others
 ORIGIN
 Query Match 91.6%; Score 17.4; DB 17; Length 672;
 Best Local Similarity 94.7%; Pred. No. 2.8e+03;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
 |||||
 Db 354 ACAAGATACAGAACATGC 372

RESULT 6
 A1383272 235 bp mRNA linear EST 18-MAR-1999
 LOCUS tc77c07.x1 Soares_NHMP1.S1 Homo sapiens cDNA clone IMAGE:2070636
 DEFINITION 3' similar to gb:662505 c5a AMAPHYLATOXIN CHEMOTACTIC RECEPTOR
 (HUMAN); mRNA sequence.
 ACCESSION A1383272
 VERSION A1383272.1 GI:4196053
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 235)
 NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 JOURNAL Unpublished (1997)
 COMMENT Contact: Robert Strausberg, Ph.D.

```

Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1836 Std Error: 0.00
Seq primer: -400P from Gidco
High quality sequence stop: 227.
Location/Qualifiers
FEATURES
SOURCE
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2070636"
/clone_lib="Soares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pUT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI: Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbH, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of 1.0A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      82 a      53 c      32 g      68 t
ORIGIN
Query Match      89.5%; Score 17; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 3.6e-03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ACAGATACAGACACAT 17
        |||||||||||||||
DB      160 ACAGATACAGACACAT 176

RESULT 7
LOCUS      BE581326      561 bp      mRNA      linear      EST 09-MAY-2001
DEFINITION      K949B04.Y1 TB955TM-SSR Strongyloides stercoralis cDNA 5' similar to
ACCESSION      BE581326
VERSION      BE581326.1 GI:9832258
KEYWORDS      EST.
SOURCE      Strongyloides stercoralis.
ORGANISM      Strongyloides stercoralis.
REFERENCE      1 (bases 1 to 561)
AUTHORS      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,T., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
JOURNAL      The Washington Univ. Nematode EST Project, 1999
COMMENT      Unpublished (1999)
CONTACT      McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIHID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 412.
Location/Qualifiers
FEATURES
SOURCE
1..561

Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1836 Std Error: 0.00
Seq primer: -400P from Gidco
High quality sequence stop: 227.
Location/Qualifiers
FEATURES
SOURCE
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2070636"
/clone_lib="Soares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pUT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI: Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbH, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of 1.0A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      82 a      53 c      32 g      68 t
ORIGIN
Query Match      89.5%; Score 17; DB 9; Length 235;
Best Local Similarity 100.0%; Pred. No. 3.6e-03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1 ACAGATACAGACACAT 17
        |||||||||||||||
DB      160 ACAGATACAGACACAT 176

RESULT 7
LOCUS      BE581326      561 bp      mRNA      linear      EST 09-MAY-2001
DEFINITION      K949B04.Y1 TB955TM-SSR Strongyloides stercoralis cDNA 5' similar to
ACCESSION      BE581326
VERSION      BE581326.1 GI:9832258
KEYWORDS      EST.
SOURCE      Strongyloides stercoralis.
ORGANISM      Strongyloides stercoralis.
REFERENCE      1 (bases 1 to 561)
AUTHORS      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,T., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
JOURNAL      The Washington Univ. Nematode EST Project, 1999
COMMENT      Unpublished (1999)
CONTACT      McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIHID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 412.
Location/Qualifiers
FEATURES
SOURCE
1..561

Email: cgabs-r@mail.nih.gov
This clone is available royalty-free through LBNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
Insert Length: 1836 Std Error: 0.00
Seq primer: -400P from Gidco
High quality sequence stop: 227.
Location/Qualifiers
FEATURES
SOURCE
1..225
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:2070636"
/clone_lib="Soares_NbHMPu_S1"
/tissue_type="Pooled human melanocyte, fetal heart, and
pregnant uterus"
/lab_host="DH10B"
/note="Organ: mixed (see below); Vector: pUT73D-Pac
(Pharmacia) with a modified polylinker; Site_1: Not 1;
Site_2: Eco RI: Equal amounts of plasmid DNA from three
normalized libraries (melanocyte 2NbH, pregnant uterus
NbHPU, and fetal heart NbH19W) were mixed, and ss circles
were made in vitro. Following HAP purification, this DNA
was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from pools of
5,000 clones made from the same 3 libraries. The pools
consisted of 1.0A.G.E. clones 260232-265223,
340488-345479, and 484488-489479."

BASE COUNT      82 a      53 c      32 g      68 t
ORIGIN
Query Match      89.5%; Score 17; DB 10; Length 561;
Best Local Similarity 100.0%; Pred. No. 4e-03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      2 CAAGATACAGACACATG 18
        |||||||||||||||
DB      394 CAAGATACAGACACATG 410

RESULT 8
LOCUS      BG226225      580 bp      mRNA      linear      EST 09-MAY-2001
DEFINITION      K918H12.Y1 TB955TM-SSR Strongyloides stercoralis cDNA 5' similar to
ACCESSION      BG226225
VERSION      BG226225.1 GI:12713780
KEYWORDS      EST.
SOURCE      Strongyloides stercoralis.
ORGANISM      Strongyloides stercoralis.
REFERENCE      1 (bases 1 to 580)
AUTHORS      McCarter,J., Clifton,S., Chiapelli,B., Pape,D., Martin,J., Wylie,T.,
Dante,M., Marra,M., Hillier,L., Kucaba,T., Theising,B., Bowers,Y.,
Gibbons,M., Ritter,E., Bennett,J., Franklin,C., Tsagarisvilli,R.,
Ronko,I., Kennedy,S., Maguire,L., Beck,C., Underwood,K., Steptoe
,M., Allen,M., Person,B., Swaller,T., Harvey,N., Schurk,R., Kohn,S.,
Shin,T., Jackson,T., Cardenas,M., McCann,R., Waterston,R. and
Wilson,R.
JOURNAL      The Washington Univ. Nematode EST Project, 1999
COMMENT      Unpublished (1999)
CONTACT      McCarter JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIHID, NIH (tnutman@nih.gov). DNA Sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 413.
Location/Qualifiers
FEATURES
SOURCE
1..580
/organism="Strongyloides stercoralis"
/strain="Rhabditiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TB955TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)
/note="Vector: Lambda uni-ZAP XR (Stratagene); Site_1:
EcoRI; Site_2: XhoI; mRNA was purified from 2 x 10E3
rhabditiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and, using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1 x 10E5 pfu/ml and an
amplified, undiluted titer of 9 x 10E11 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."

```


and, using adaptors, was cloned unidirectionally into the vector from the EcoRI site to the XhoI site. The library has an unamplified titer of 1×10^5 pfu/ml and an amplified, undiluted titer of 1×10^8 pfu/ml. The average insert size of the unamplified library is 675 bp (range, 100-1700)."

BASE COUNT 206 a 99 c 115 g 160 t
ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 580;
Best Local Similarity 100.0%; Pred. No. 4e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CAAGATACAGACAAATG 18
|||||
Db 386 CAAGATACAGACAAATG 402

RESULT 9
LOCUS BG227716 601 bp mRNA linear EST 09-MAY-2001
DEFINITION kq12c01.y1 TB95TM-SSR Strongyloides stercoralis cDNA 5' similar to
ACCESSION BG227716
VERSION BG227716.1 GI:12715200
KEYWORDS EST.
SOURCE Strongyloides stercoralis.
ORGANISM Strongyloides stercoralis
Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida;
Panagrolaimoidea; Strongyloidea; Strongyloides.
REFERENCE
1 (bases 1 to 601)
McCarte, J., Clifton, S., Chapelli, B., Pape, D., Martin, J., Wylie, T.,
Dante, M., Marra, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, V.,
Ronko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R., and
Wilson, R.
TITLE The Washington Univ. Nematode EST Project, 1999
JOURNAL Unpublished (1999)
COMMENT Contact: McCarte JP
The Washington Univ. Nematode EST Project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wustl.edu
The library was constructed by Dr. Thomas Nutman and colleagues of
NIH (nutman@nih.gov). DNA sequencing by: Washington
University Genome Sequencing Center St. Louis.
High quality sequence stop: 414.
Location/Qualifiers
1. 601

FEATURES
SOURCE
/organism="Strongyloides stercoralis"
/strain="Rhabdilitiform larvae obtained from gerbils"
/db_xref="taxon:6248"
/clone_lib="TB95TM-SSR"
/lab_host="XL-1 Blue MRF" (Stratagene)"
/note="Vector: Lambda Uni-ZAP XR (Stratagene). Site 1:
EcoRI; Site 2: XhoI; mRNA was purified from 2 x 10^5
rhabdilitiform larvae which had been isolated from gerbils
experimentally infected with larvae originally isolated
from experimentally infected dogs. cDNA was constructed
and, using adaptors, was cloned unidirectionally into the
vector from the EcoRI site to the XhoI site. The library
has an unamplified titer of 1×10^5 pfu/ml and an
amplified, undiluted titer of 1×10^8 pfu/ml. The
average insert size of the unamplified library is 675 bp
(range, 100-1700)."

BASE COUNT 213 a 102 c 120 g 163 t 3 others
ORIGIN

Query Match 89.5%; Score 17; DB 12; Length 601;

Best Local Similarity 100.0%; Pred. No. 4.1e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 CAAGATACAGACAAATG 18
|||||
Db 386 CAAGATACAGACAAATG 402

RESULT 10
LOCUS BH202699 721 bp DNA linear GSS 24-OCT-2001
DEFINITION Sml-44M4.TR Sml Schistosoma mansoni genomic clone Sml-44M4, DNA
sequence.
ACCESSION BH202699
VERSION BH202699.1 GI:16375476
KEYWORDS GSS.
SOURCE Schistosoma mansoni.
ORGANISM Schistosoma mansoni.
Schistosoma mansoni.
Eukaryota; Metazoa; Platyhelminthes; Trematoda; Digenea;
Strigeida; Schistosomatoidea; Schistosomatidae; Schistosoma.
REFERENCE
1 (bases 1 to 721)
Shetty, J., Simpson, A., Malek, J., Koo, H., LovVerde, P. T. and El-Sayed,
N. M.
Use of end sequences from Schistosoma mansoni (Puerto Rico strain)
Sml BAC library for gene discovery and map construction
Unpublished (2001)
Other GSSs: Sml-44M4.TF
Contact: Najib M. El-Sayed
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: nelsayed@igr.org
Lo-edu).
Seq primer: M13 Rev
Class: BAC ends.
Location/Qualifiers
1. 721
/organism="Schistosoma mansoni"
/strain="Puerto Rico"
/db_xref="taxon:6183"
/clone_lib="Sml-44M4"
/note="Vector: pBelOBAC11; Site 1: Hin dIII; Constructed
in the laboratory of Dr. Denis Le Paslier at the Fondation
Jean Dausset, CEPH, Paris, France. Briefly, Schistosoma
mansoni agarose embedded DNA was partially digested with
Hin dIII. High molecular weight fragments were ligated in
pBelOBAC11 digested with Hin dIII. The average insert size
is 100 kb. Total clone coverage: approx. 7.95 x the
haploid genome. Further information can be found in Le
Paslier et al. (2000) Construction and characterization of
a Schistosoma mansoni bacterial artificial chromosome
library. Genomics 65: 87-94."

BASE COUNT 222 a 151 c 136 g 212 t
ORIGIN

Query Match 89.5%; Score 17; DB 17; Length 721;
Best Local Similarity 100.0%; Pred. No. 4.2e+03;
Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACAAAT 17
|||||
Db 562 ACAAGATACAGACAAAT 578

RESULT 11
LOCUS AM749252 149 bp mRNA linear EST 28-APR-2000
DEFINITION CM2-BT0364-271299-061-e02 BT0364 Homo sapiens cDNA, mRNA sequence.
ACCESSION AM749252
VERSION AM749252.1 GI:7664184

KEYWORDS	EST.	human.
SOURCE		
ORGANISM	Homo sapiens	
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 149)	
AUTHORS	Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bala, G. S., Simpson, D. H., Brunstein, A. A., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.	
TITLE	Shoyun sequence of the human transcriptome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&tl2=CM2-BT0364-1) 271299-061-a02&f3=1999-12-27&td=1) Seq primer: puc 18 forward High quality sequence start: 75 High quality sequence start: 115. Location/Qualifiers 1. 149 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_1lb="BT0364" /der="Adult" /note="Organ: Breast; Vector: puc18; Site.1: SnaI; Site.2: SnaI; Site.3: SmaI; A mini-library was made by cloning products derived from ORSTS PCR (U.S. Letters Patent application No. 1967,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	51 a 28 c 35 g 35 t	
ORIGIN		
Query Match	86.38;	Score 16.4; DB 10; Length 149;
Best Local Similarity	94.44;	Pred. No. 5; 9e+03;
Matches 17; Conservative 0;	Mismatches 1;	Indels 0; Gaps 0;
Oy	2 CAGATTCAGACGATGC 19	
Db	52 CAGATTCAGATCATGC 69	
RESULT 12		
AA742620		
LOCUS	AA742620	213 bp mRNA linear EST 23-JAN-1998
DEFINITION	my90b11.s1 NCI CGAP, GCD1 Homo sapiens cDNA IMAGE:1285533 3'	
	similar to gp:MI0901 GUDCOCORTICOID RECEPTOR, ALPHA (HUMAN);, mRNA	
ACCESSION	AA742620	
VERSION	AA742620.1	GI:2784441
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. 1 (bases 1 to 213)	
AUTHORS	NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	

JOURNAL
 COMMENT
 Unpublished (1997)
 Contact: Robert Strassberg, Ph.D.
 Email: c9apb5-remail.nih.gov
 Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.
 cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 cDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.lnln.gov/dbfp/image/image.html

FEATURES
 SOURCE
 Trace considered overall poor quality
 Insert Length: 1113 Std Error: 0.00
 Seq primer: -40ml3 fwd. E7 from Amersham
 High quality sequence stop: 1.
 Location/Qualifiers
 1..213
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:128533"
 /clone_1b="NCI-CGAP-CGB1"
 /rissc_type="germinal center B cell"
 /lab_host="DH10B"
 /note="Vector: pT73D-Pac (pharmacia) with a modified polynker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - 0190(dT) primer
 15'-GTGACCATCTGTAAGTGGACGCGCGCTCATTTTCTTTTCTT-3'
]. Double-stranded cDNA was ligated to Eco RI adaptors (pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT
 ORIGIN
 75 a 31 c 37 g 70 t

Query Match 86.3%; Score 16.4; DB 9; Length 213;
 Best Local Similarity 94.4%; Pred. No. 6.2e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 2 CAAGTACAGACATGC 19
 Db 38 CAAGTACTGACACATGC 55
 ||||||| |||||||

RESULT 13
 T97764/c 261 bp mRNA linear EST 29-MAR-1995
 LOCUS T97764
 DEFINITION IMAGE:121954 3', mRNA sequence.
 VERSION T97764
 ACCESSION T97764.1 GI:747109
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 REFERENCE 1 (bases 1 to 261)
 AUTHORS Haller, L., Clark, N., Dubuque, T., Ellison, K., Hawkins, M., Holman, M., Holman, M., Kucaba, T., Le, M., Lennon, O., Matis, M., Parsons, J., Rifkin, L., Roilting, T., Soares, M., Tan, F., Tevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and Wilson, R.
 TITLE The WashU-Merck EST Project
 JOURNAL Unpublished (1995)
 COMMENT Contact: Wilson RK
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800

KEYWORDS	EST.	human.
SOURCE		
ORGANISM	Homo sapiens	
REFERENCE	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
AUTHORS	1 (bases 1 to 149) Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M. R., Nagal, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Bata, G. S., Simpson, D. H., Brunstein, A. E., de Oliveira, P. S., Bucher, P., Jongeneel, C. V., O'Hare M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J. J.	
TITLE	Shoiguon sequencing of the human transcriptome with ORF expressed sequence tags	
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	
MEDLINE	20202663	
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM2&ct=CM2-BT0364 271289-061-a02&ct3=1999-12-27&td=1) Seq primer: puc 18 forward High quality sequence start: 75 High quality sequence start: 115. Location/Qualifiers 1..149 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_lib="H0364" /dev_stage="Adult" /note="Organ: Breast; Vector: puc18; Site.1: SmaI; Site.2: SmaI; SmaI: A mini-library was made by cloning products derived from ORSSTPS PCR (U.S. Letters Patent application No. 196 716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."	
BASE COUNT	51 a 28 c 35 g 35 t	
ORIGIN		
Query Match	86.3%; Score 16.4; DB 10; Length 149;	
Best Local Similarity	94.4%; Pred. No. 5.9e-03;	
Matches	17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	
Oy	2 CAGATTCAGACACATGC 19 	
Db	52 CAGATTCAGACATATGC 69 	
RESULT 12		
LOCUS	AA742620 213 bp mRNA linear EST 23-JAN-1998	
DEFINITION	np04011.s1 NCI CGAP GCBI Homo sapiens cDNA IMAGE:128553 3'	
ACCESSION	AA742620	
VERSION	AA742620	
KEYWORDS	sequence.	
SOURCE	AA742620.1 GI:2784441	
ORGANISM	EST. human.	
REFERENCE	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.	
AUTHORS	1 (bases 1 to 213) NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.	
TITLE	National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index	

JOURNAL

Unpublished (1997)

Contact: Robert Strassberg, Ph.D.

Email: cgsab@remail.nih.gov

Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at:

www.bio.lnlnl.gov/dbfp/image/image.html

FEATURES

SOURCE

Trace considered overall poor quality

Insert Length: 1113 Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amerham

High quality sequence stop: 1.

Location/Qualifiers

1..213

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:128533"

/clone_1b="NCI_CGAP-GCB1"

/rname_type="germinal center B cell"

/lab_host="DH10B"

/note="Vector: pT73D-Pac (pharmacia) with a modified polynker; Site.1: Not I; Site.2: Eco RI; 1st strand cDNA prepared from human tonsillar cells enriched for germinal center B cells by flow sorting (CD20+, IgD-), provided by Dr. Louis M. Staudt (NCI), Dr. David Allman (NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was primed with a Not I - o19o(dT) primer

15'-GTGACATCTCAAGTCGGACCGCCGCATCTTTTCTTTTCTT-3'

]. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT

75 a 31 c 37 g 70 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 9; Length 213;

Best Local Similarity 94.4%; Pred. No. 6.2e+03;

Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CAAGTACAGACATGC 19

I|||||I|||||I|||||I

Db 38 CAAGTACTGACAAATGC 55

RESULT 13

T97764/c 261 bp mRNA linear EST 29-MAR-1995

LOCUS T97764.c

DEFINITION Y68606.s1 Soares fetal liver spleen INLIS Homo sapiens cDNA clone IMAGE:121954 3', mRNA sequence.

VERSION T97764

ACCESSION T97764.1 GI:747109

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

REFERENCE 1 (bases 1 to 261)

Hallier,L., Clark,N., Dubugue,T., Ellison,K., Hawkins,M., Holman,M., Hutman,M., Kucaba,T., Le,M., Lennon,O., Marks,W., Parsons,J., Rifkin,H., Roilting,T., Soares,M., Tan,F., Trevasakis,E., Waterston,R., Williamson,A., Wohlmann,P. and Wilson,R.

The WashU-Merck EST Project

Unpublished (1995)

CONTACT: Wilson RK

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810
Email: estewatson.wustl.edu
Insert Size: 977
High quality sequence stops: 256 Source: IMAGE Consortium, LLM
This clone is available royalty-free through LLM; contact the
IMAGE Consortium (infoimage.llnl.gov) for further information.
Possible reversed clone: polyT not found
Insert Length: 977 Std Error: 0.00
Seq primer: -21m13
High quality sequence stop: 256.

FEATURES
source
location/Qualifiers
1. 261
/organism="Homo sapiens"
/db_xref="GDB:474499"
/db_xref="taxon:9606"
/clone="IMAGE:121954"
/clone_lib="Soares fetal liver spleen INPIS"
/sex="male"
/dev_stage="20 week-post conception fetus"
/lab_host="DH10B (ampicillin resistant)"
/note="Organ: Liver and Spleen; Vector: pT73D (Pharmacia)
with a modified polylinker; Site 1: Pac I; Site 2: Eco RI;
1st strand cDNA was primed with a Pac I - oligo(dT) primer
[5' AACTGGAAGATTATTAAGATCTTTTCTTTTCTTTT 3'],
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Pac I and cloned into the Pac I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization. Library
constructed by Bento Soares and M. Patricia Bonaldi."

BASE COUNT
ORIGIN
90 a 36 c 57 g 77 t 1 others

Query Match
Best Local Similarity 86.3%; Score 16.4; DB 14; Length 261;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 171 CAAGATACGACGACATTCG 154
|||||
|||||

RESULT 14
BE044749
LOCUS
DEFINITION
hm30a05.X1 NCI_CGAP_Thy7 Homo sapiens cDNA clone IMAGE:3023600 3',
mRNA sequence.
ACCESSION
BE044749
VERSION
BE044749.1 GI:8361802
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
1 (bases 1 to 309)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
AUTHORS
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
TITLE
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapsr@mail.nih.gov
Tissue Procurement: Klaus Kaserer M.D., Bruno Niederle M.D., Mike
Emmert-Buck M.D., Ph.D., Vlado Knezevic M.D.
cDNA Library Preparation: Krizman Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Possible reversed clone: polyT not found
Seq primer: -400P from Gibco
High quality sequence stop: 308.
location/Qualifiers
1. 309
source

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3023600"
/clone_lib="NCI_CGAP_Thy7"
/tissue_type="follicular adenoma (benign lesion)"
/lab_host="DH10B"
/note="Organ: thyroid; Vector: pAMP10; mRNA made from
thyroid adenoma (benign), cDNA made by oligo-dT priming.
Non-directionally cloned into UBC sites. Size-selected on
agarose gel, average insert size 500 bp. Primary library.
cDNA Library Preparation: David B. Krizman, Ph.D.
REFERENCE: Krizman et al. (1996) Cancer Research
56:5380-5383."

BASE COUNT
ORIGIN
87 a 52 c 55 g 115 t

Query Match
Best Local Similarity 86.3%; Score 16.4; DB 10; Length 309;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 203 ACAAGATACGACGACGACG 220
|||||
|||||

RESULT 15
BH269502
LOCUS
DEFINITION
CH230-82021.1V CHORI-230 Segment 1 Rattus norvegicus genomic clone
CH230-82021. DNA sequence.
ACCESSION
BH269502
VERSION
BH269502.1 GI:17181808
KEYWORDS
GSS.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 309)
Zhao, S., Shetty, J., Shatsman, S., Tsegaye, G., Geer, K., Shvartbeyn
A., Gebregorgis, E., Overton, L., Russell, D., Chen, D., Riggs, F., de
Jong, P. and Fraser, C.M.
Rat BAC End Sequences from Library CHORI-230 EcORI segment
Unpublished (1999)
CONTACT: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genome Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the rat BAC library CHORI-230
(<http://www.chori.org/bacpac/rat230.htm>). For BAC library
availability, please contact Pieter de Jong (pdejong@mail.cho.org).
Clones may be purchased from BACPAC Resources
(http://www.chori.org/bacpac/or_eiting_information.htm). BAC end
page: http://www.tigr.org/tidb/bac_ends/rat/bac_end_intro.html
Plate: 82 row: 0 column: 21
Seq primer: 77
Class: BAC ends.
location/Qualifiers
1. 309
/organism="Rattus norvegicus"
/strain="BN/SsNhsd/MCw"
/db_xref="taxon:10116"
/clone="CH230-82021"
/clone_lib="CHORI-230 Segment 1"
/sex="Female"
/cell_type="Brain"
/note="Vector: pFARAC2.1; Site 1: EcoRI; Site 2: EcoRI;
CHORI-230 Rat (BN/SsNhsd/MCw) BAC library produced by
Pieter de Jong"

BASE COUNT
95 a 57 c 54 g 103 t

ORIGIN

Query Match 86.3%; Score 16.4; DB 17; Length 309;
 Best Local Similarity 94.4%; Pred. No. 6.5e+03;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAAGATCAGACAAATGC 19
 ||||| |||||
 DB 11 CAAGATCAGACAAATGC 28

Search completed: January 23, 2003, 22:08:56
 Job time : 1384.4 secs

Query Match 77.9%; Score 14.8; DB 5; Length 3725;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAAGATACAGAACATGC 18
|||||
DB 2300 CAAGATACAGAACATGC 2317

RESULT 14

US-08-095-737-1
Sequence 1, Application US/08095737
Patent No. 5487979
GENERAL INFORMATION:
APPLICANT: DiFiore, Pier P
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: A Substrate for the Epidermal Growth
TITLE OF INVENTION: Factor Receptor Kinase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/095,737
FILING DATE: 19930722
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 21..2709
US-08-095-737-1
Query Match 77.9%; Score 14.8; DB 1; Length 4165;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CAAGATACAGAACATGC 19
|||||
DB 4044 CAAGATACAGAACATGC 4061

TITLE OF INVENTION: Factor Receptor Kinase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSER: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, Sixteenth Floor
CITY: Newport Beach
STATE: California
COUNTRY: United States of America
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,145
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/095,737
FILING DATE: 22-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Israelson, Ned A
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH060.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4165 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FEATURE:
NAME/KEY: CDS
LOCATION: 21..2709
US-08-480-145-1
Query Match 77.9%; Score 14.8; DB 1; Length 4165;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 2 CAAGATACAGAACATGC 19
|||||
DB 4044 CAAGATACAGAACATGC 4061

Search completed: January 23, 2003, 20:35:59
Job time : 35.6571 secs

REGISTRATION NUMBER: 31-684
REFERENCE/DOCKET NUMBER: 92-21C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3725 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APPH
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2364
US-08-424-022-12

Query Match 77.9%; Score 14.8; DB 1; Length 3725;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGAACATG 18
||||| ||||| |||
DB 2300 ACAAGATGACAGAACATG 2317

RESULT 12
US-08-424-017B-12
Sequence 12, Application US/08424017B
Patent No. 5935854
GENERAL INFORMATION:
APPLICANT: Foster, Donald C
APPLICANT: Sprecher, Cindy
TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR
TITLE OF INVENTION: HOWOLOG AND KUNITZ-TYPE INHIBITOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,017B
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/155,331
FILING DATE:
APPLICATION NUMBER: US 07/985,692
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-684
REFERENCE/DOCKET NUMBER: 92-21C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3725 base pairs
TYPE: nucleic acid

STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APPH
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2364
US-08-424-017B-12

Query Match 77.9%; Score 14.8; DB 2; Length 3725;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGAACATG 18
||||| ||||| |||
DB 2300 ACAAGATGACAGAACATG 2317

RESULT 13
PCT-US93-11696-12
Sequence 12, Application PC/TUS9311696
GENERAL INFORMATION:
APPLICANT: Foster, Donald C
APPLICANT: Sprecher, Cindy
TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR
TITLE OF INVENTION: HOWOLOG AND KUNITZ-TYPE INHIBITOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11696
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,692
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-684
REFERENCE/DOCKET NUMBER: 92-21C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3725 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APPH
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2364
PCT-US93-11696-12

INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 693 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..693
US-08-168-091A-3

Query Match 78.9%; Score 15; DB 1; Length 693;
Best Local Similarity 100.0%; Pred. No. 1.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CAAGATACAGACAA 16
DB 150 CAAGATACAGACAA 164

RESULT 9
US-09-134-001C-470
Sequence 470, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GYC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
PRIOR FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 470
LENGTH: 1521
TYPE: DNA
ORGANISM: Staphylococcus epidermidis
US-09-134-001C-470

Query Match 77.9%; Score 14.8; DB 4; Length 1521;
Best Local Similarity 88.9%; Pred. No. 1.8e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGACATG 18
DB 489 ACAAGATACAGACATG 506

RESULT 10
US-08-155-331-12
Sequence 12, Application US/08155331
Patent No. 5441931
GENERAL INFORMATION:
APPLICANT: Foster, Donald C
APPLICANT: Sprecher, Cindy
APPLICANT: No. 5441931-15, Kjeld
TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR
TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/155,331
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,692
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E
REGISTRATION NUMBER: 31-684
REFERENCE/DOCKET NUMBER: 92-21C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-547-8080 ext 322
TELEFAX: 206-548-2329
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 3725 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE: Homo sapiens
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APFH
FEATURE:
NAME/KEY: CDS
LOCATION: 73..2364
US-08-155-331-12

Query Match 77.9%; Score 14.8; DB 1; Length 3725;
Best Local Similarity 88.9%; Pred. No. 1.9e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGACATG 18
DB 2300 ACAAGATACAGACATG 2317

RESULT 11
US-08-424-022-12
Sequence 12, Application US/08424022
Patent No. 5677146
GENERAL INFORMATION:
APPLICANT: Foster, Donald C
APPLICANT: Sprecher, Cindy
APPLICANT: No. 5677146-15, Kjeld
TITLE OF INVENTION: NOVEL HUMAN AMYLOID PROTEIN PRECURSOR
TITLE OF INVENTION: HOMOLOG AND KUNITZ-TYPE INHIBITOR
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESSEE: ZymoGenetics, Inc.
STREET: 4225 Roosevelt Way, N.E.
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/424,022
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,692
FILING DATE: 02-DEC-1992
ATTORNEY/AGENT INFORMATION:
NAME: Parker, Gary E

GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gunney, Austin L.
APPLICANT: Hillen, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P117652
CURRENT APPLICATION NUMBER: US/09/182,145B
EARLIER FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: US 60/063,704
EARLIER FILING DATE: 1997-10-29
EARLIER APPLICATION NUMBER: US 60/073,612
EARLIER FILING DATE: 1998-02-04
EARLIER APPLICATION NUMBER: US 60/081,695
EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 31
LENGTH: 1335
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Unsure
LOCATION: 18, 131
OTHER INFORMATION: Unknown base.
US-09-182-145-31

Query Match 83.2%; Score 15.8; DB 4; Length 1335;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATCAGACAAATGC 19
||||| ||||| ||||| |||||
Db 669 ACAAGCTACAAACATGC 651

RESULT 7
US-08-685-576-5
Sequence 5, Application US/08685576
Patent No. 5906819
GENERAL INFORMATION:
APPLICANT: Kaibuchi, Koro
APPLICANT: Iwamatsu, Akihito
APPLICANT: Nakano, Takeshi
APPLICANT: Ito, Masaaki
APPLICANT: Takahashi, No. 5906819uaki
TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/685,576
FILING DATE: 24-JUL-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 7-325129
FILING DATE: 20-NOV-1995
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-17150
FILING DATE: 05-JAN-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 8-131206
FILING DATE: 26-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 16887/843
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4363 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 1..4164
US-08-685-576-5

Query Match 83.2%; Score 15.8; DB 2; Length 4363;
Best Local Similarity 89.5%; Pred. No. 68;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATCAGACAAATGC 19
||||| ||||| ||||| |||||
Db 3564 ACAAGATAAGAACATGC 3582

RESULT 8
US-08-168-091A-3
Sequence 3, Application US/08168091A
Patent No. 5665862
GENERAL INFORMATION:
APPLICANT: Fischbach, Gerald.
APPLICANT: Falls, Douglas R.
APPLICANT: Rosen, Kenneth M.
APPLICANT: Corfas, Gabriel
TITLE OF INVENTION: Neurotrophic Factor
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSER: LARIVE AND COCKFIELD
STREET: 60 State Street, Suite 510
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: ASCII text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/168,091A
FILING DATE: 15-DEC-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/953,742
FILING DATE: 29-SEP-1992
ATTORNEY/AGENT INFORMATION:
NAME: DeConti, Giulio A.
REGISTRATION NUMBER: 31,503
REFERENCE/DOCKET NUMBER: HMT-002CP
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 227-7400
TELEFAX: (617) 227-5941

ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (17)...(1078)
US-09-253-316-1

Query Match 83.2%; Score 15.8; DB 4; Length 1142;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
||||| ||||| ||||| |||||
DB 584 ACAAGCTCAAAACATGC 602

RESULT 3

US-09-182-145-34
Sequence 34, Application US/09182145B
Patent No. 6387657
GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P1176R2
CURRENT APPLICATION NUMBER: US/09/182,145B
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: US 60/063,704
EARLIER FILING DATE: 1997-10-29
EARLIER APPLICATION NUMBER: US 60/073,612
EARLIER FILING DATE: 1998-02-04
EARLIER APPLICATION NUMBER: US 60/081,695
EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 34
LENGTH: 1212
TYPE: DNA
ORGANISM: Homo sapiens
US-09-182-145-34

Query Match 83.2%; Score 15.8; DB 4; Length 1212;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
||||| ||||| ||||| |||||
DB 583 ACAAGCTCAAAACATGC 601

RESULT 4

US-09-182-145-35/C
Sequence 35, Application US/09182145B
Patent No. 6387657
GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P1176R2

CURRENT APPLICATION NUMBER: US/09/182,145B
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: US 60/063,704
EARLIER FILING DATE: 1997-10-29
EARLIER APPLICATION NUMBER: US 60/073,612
EARLIER FILING DATE: 1998-02-04
EARLIER APPLICATION NUMBER: US 60/081,695
EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 35
LENGTH: 1212
TYPE: DNA
ORGANISM: Homo sapiens
US-09-182-145-35

Query Match 83.2%; Score 15.8; DB 4; Length 1212;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
||||| ||||| ||||| |||||
DB 630 ACAAGCTCAAAACATGC 612

RESULT 5

US-09-182-145-30
Sequence 30, Application US/09182145B
Patent No. 6387657
GENERAL INFORMATION:
APPLICANT: Botstein, David A.
APPLICANT: Cohen, Robert
APPLICANT: Goddard, Audrey
APPLICANT: Gurney, Austin L.
APPLICANT: Hillan, Kenneth J.
APPLICANT: Lawrence, David A.
APPLICANT: Levine, Arnold J.
APPLICANT: Pennica, Diane
APPLICANT: Roy, Margaret Ann
APPLICANT: Wood, William I.
TITLE OF INVENTION: WISP POLYPEPTIDES AND NUCLEIC ACIDS ENCODING SAME
FILE REFERENCE: P1176R2
CURRENT APPLICATION NUMBER: US/09/182,145B
CURRENT FILING DATE: 1998-10-29
EARLIER APPLICATION NUMBER: US 60/063,704
EARLIER FILING DATE: 1997-10-29
EARLIER APPLICATION NUMBER: US 60/073,612
EARLIER FILING DATE: 1998-02-04
EARLIER APPLICATION NUMBER: US 60/081,695
EARLIER FILING DATE: 1998-04-14
NUMBER OF SEQ ID NOS: 156
SEQ ID NO 30
LENGTH: 1335
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: Unsure
LOCATION: 1205, 1318
OTHER INFORMATION: Unknown base.
US-09-182-145-30

Query Match 83.2%; Score 15.8; DB 4; Length 1335;
Best Local Similarity 89.5%; Pred. No. 63;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGAACATGC 19
||||| ||||| ||||| |||||
DB 667 ACAAGCTCAAAACATGC 685

RESULT 6
US-09-182-145-31/C
Sequence 31, Application US/09182145B
Patent No. 6387657

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 19:22:47 ; Search time 33 6571 Seconds
(without alignments)
173.124 Million cell updates/sec

Title: US-09-803-736-1079_COPY_13662_13680

Sequence: 19
1 acaagatcacagacatgc 19

Scoring table: IDENTITY_NUC
Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/2/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PTCUS.COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfillseq1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
c 1	15.8	83.2	33	4	US-09-253-316-17
2	15.8	83.2	1142	4	US-09-253-316-1
3	15.8	83.2	1212	4	US-09-182-145-34
c 4	15.8	83.2	1212	4	US-09-182-145-35
5	15.8	83.2	1335	4	US-09-182-145-30
c 6	15.8	83.2	1335	4	US-09-182-145-31
7	15.8	83.2	4363	2	US-08-685-576-5
8	15.8	78.9	693	1	US-08-168-091A-3
9	14.8	77.9	1521	4	US-09-134-001C-470
10	14.8	77.9	3725	1	US-08-155-331-12
11	14.8	77.9	3725	1	US-08-424-022-12
12	14.8	77.9	3725	2	US-08-424-017B-12
13	14.8	77.9	3725	5	PCT-US93-11696-12
14	14.8	77.9	4165	1	US-08-095-737-1
15	14.8	77.9	4165	1	US-08-480-145-1
16	14.8	77.9	4165	2	US-08-477-389-1
17	14.8	77.9	5690	2	US-08-447-464-2
18	14.8	77.9	5690	2	US-08-716-679-2
19	14.8	77.9	6000	1	US-08-348-006B-6
20	14.8	77.9	6000	2	US-08-800-825A-6
21	14.8	77.9	6000	4	US-09-158-657-6
22	14.8	77.9	6000	5	PCT-US94-10166-6
c 23	14.8	77.9	7760	4	US-08-961-527-63
c 24	14.8	77.9	11236	1	US-07-853-913-1
25	14.4	75.8	763	4	US-09-063-733A-12
c 26	14.4	75.8	1703	1	US-07-868-353A-11
c 27	14.4	75.8	1703	1	US-08-407-804-20

c 28	14.4	75.8	1703	3	US-09-124-807-20	Sequence 20, Appl
c 29	14.4	75.8	1734	4	US-09-146-969-5	Sequence 5, Appl
c 30	14.4	75.8	1782	4	US-09-604-978-2	Sequence 2, Appl
31	14.4	75.8	1854	4	US-09-314-242-1	Sequence 1, Appl
32	14.4	75.8	1883	4	US-09-063-733A-45	Sequence 45, Appl
33	14.4	75.8	1950	4	US-09-063-733A-50	Sequence 50, Appl
34	14.4	75.8	2512	4	US-09-248-757-1	Sequence 1, Appl
35	14.4	75.8	2700	4	US-09-063-733A-15	Sequence 15, Appl
36	14.4	75.8	3572	4	US-09-575-574-3	Sequence 3, Appl
37	14.4	75.8	5053	2	US-08-685-576-2	Sequence 3, Appl
c 38	14.4	75.8	13104	4	US-08-961-527-34	Sequence 34, Appl
39	14.4	75.8	15378	3	US-08-785-420-1	Sequence 1, Appl
40	14.2	74.7	783	5	PCT-US92-00282-22	Sequence 22, Appl
c 41	14.2	74.7	896	4	US-09-171-209-5	Sequence 5, Appl
42	14.2	74.7	3870	1	US-08-138-641-1	Sequence 1, Appl
43	14.2	74.7	3870	1	US-08-138-133-1	Sequence 1, Appl
44	14.2	74.7	3893	1	US-08-138-641-3	Sequence 3, Appl
45	14.2	74.7	3893	1	US-08-138-133-3	Sequence 3, Appl

ALIGNMENTS

```

RESULT 1
US-09-253-316-17/c
; Sequence 17, Application US/09253316
; Patent No. 635890
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR HOMOLOGS
; FILE REFERENCE: 97-75
; CURRENT APPLICATION NUMBER: US/09/253,316
; CURRENT FILING DATE: 1999-02-19
; EARLIER APPLICATION NUMBER: US 60/075,300
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 33
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: oligonucleotide primer ZC16422
US-09-253-316-17

Query Match      83.2%   Score 15.8 ; DB 4; Length 33;
Best Local Similarity 89.5%   Pred. No. 51;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATCACAGACATGC 19
    ||||| ||||| ||||| |||||
DB 30 ACAAGCTACAAAACATGC 12

RESULT 2
US-09-253-316-1
; Sequence 1, Application US/09253316
; Patent No. 635890
; GENERAL INFORMATION:
; APPLICANT: Sheppard, Paul O.
; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR HOMOLOGS
; FILE REFERENCE: 97-75
; CURRENT APPLICATION NUMBER: US/09/253,316
; CURRENT FILING DATE: 1999-02-19
; EARLIER APPLICATION NUMBER: US 60/075,300
; EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 1142
; TYPE: DNA

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 18:49:02 : Search time 35.2857 Seconds
(without alignments)
241.912 Million cell updates/sec

Title: US-09-803-736-1079_COPY_13662_13680
Perfect score: 19
Sequence: 1 acaagatacagacatgc 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 396772 seqs, 224632407 residues 793544
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA.*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/1/pubpna/PC1_NEM_PUB.seq.*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq.*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq.*
6: /cgn2_6/ptodata/1/pubpna/PC1US_PUBCOMB.seq.*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq.*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq.*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq.*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq.*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Match Length	DB ID	Description
C 1	16	84.2	6659 10	US-09-321-967B-1
C 2	15.8	83.2	33 12	US-10-011-859-17
C 3	15.8	83.2	159 10	US-09-867-701-9755
C 4	15.8	83.2	187 10	US-09-867-701-9964
5	15.8	83.2	366 10	US-09-925-297-422
6	15.8	83.2	375 9	US-09-738-626-1999
7	15.8	83.2	1142 12	US-10-011-859-1
8	15.8	83.2	6409 10	US-09-864-864-293
9	15.8	83.2	203654 10	US-09-820-905-3
C 10	15.4	81.1	475 10	US-09-770-444-72
11	15	78.9	326014 10	US-09-731-231A-3
12	14.8	77.9	175 9	US-10-046-935-1011
13	14.8	77.9	175 9	US-09-878-178-1011
C 14	14.8	77.9	242 10	US-09-898-598-1959
C 15	14.8	77.9	272 10	US-09-867-701-54
C 16	14.8	77.9	405 9	US-09-938-842A-2641
C 17	14.8	77.9	499 10	US-09-783-590-4188
18	14.8	77.9	500 10	US-09-728-446-794
19	14.8	77.9	611 10	US-09-925-302-431

20	14.8	77.9	1038 9	US-09-938-842A-3994	Sequence 3994, App
21	14.8	77.9	1040 10	US-09-887-576-696	Sequence 696, App
22	14.8	77.9	1402 10	US-09-764-869-2274	Sequence 2274, App
23	14.8	77.9	1402 10	US-09-764-869-2275	Sequence 2275, App
24	14.8	77.9	1402 10	US-09-764-869-2276	Sequence 2276, App
25	14.8	77.9	2020 10	US-09-925-302-356	Sequence 356, App
26	14.8	77.9	2113 10	US-09-739-254-48	Sequence 48, App
27	14.8	77.9	2113 10	US-09-904-613-48	Sequence 48, App
28	14.8	77.9	2287 10	US-09-925-302-88	Sequence 88, App
29	14.8	77.9	2643 10	US-09-822-849A-228	Sequence 228, App
30	14.8	77.9	3236 10	US-09-925-297-116	Sequence 116, App
31	14.8	77.9	3397 10	US-09-880-107-2403	Sequence 2403, App
32	14.8	77.9	3900 12	US-10-044-090-451	Sequence 451, App
33	14.8	77.9	3934 9	US-10-174-590-349	Sequence 349, App
34	14.8	77.9	3934 9	US-10-176-758-349	Sequence 349, App
35	14.8	77.9	3934 9	US-10-175-737-349	Sequence 349, App
36	14.8	77.9	3934 12	US-10-052-586-349	Sequence 349, App
C 37	14.8	77.9	9021 9	US-09-070-927A-155	Sequence 155, App
38	14.8	77.9	9392 9	US-09-764-868-1325	Sequence 1325, App
39	14.8	77.9	12121 9	US-09-764-868-1327	Sequence 1327, App
40	14.8	77.9	12121 9	US-09-764-868-1500	Sequence 1500, App
C 41	14.8	77.9	12126 9	US-09-764-868-1328	Sequence 1328, App
C 42	14.8	77.9	90541 10	US-09-759-359A-3	Sequence 3, App
C 43	14.8	77.9	170834 10	US-09-835-232-7	Sequence 7, App
C 44	14.4	75.8	246 10	US-09-878-574-6312	Sequence 6312, App
C 45	14.4	75.8	347 10	US-09-867-701-6654	Sequence 6654, App

ALIGNMENTS

RESULT 1
US-09-321-967B-1/c
Sequence 1, Application US/09321987B
Patent No. US2002010210A1
GENERAL INFORMATION:
APPLICANT: Bielloch, Judith E
APPLICANT: Bielloch, Robert H
TITLE OF INVENTION: Agent and Method for Modulating Cell Migration
FILE REFERENCE: 960296 95386
CURRENT FILING DATE: 1999-05-28
PRIOR APPLICATION NUMBER: 60/087,170
PRIOR FILING DATE: 1998-05-29
PRIOR APPLICATION NUMBER: 60/129,023
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 1
LENGTH: 6659
TYPE: DNA
ORGANISM: Caenorhabditis elegans
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(6450)
US-09-321-967B-1
Query Match 84.2%; Score 16; DB 10; length 6659;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0;
Gaps 0;
C 1 ACAAGATACAGAACAA 16
Db 3723 ACAAGATACAGAACAA 3708
RESULT 2
US-10-011-859-17/c
Sequence 17, Application US/10011859
Patent No. US20020147328A1
GENERAL INFORMATION:
APPLICANT: Sheppard, Paul O.
APPLICANT: Jaspers, Stephen R.

```

; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR HOMOLOGS
; FILE REFERENCE: 97-75
; CURRENT APPLICATION NUMBER: US/10/011,859
; CURRENT FILING DATE: 2001-11-05
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/253,316
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-19
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/075,300
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-20
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 17
; LENGTH: 33
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: oligonucleotide primer zc16422
US-10-011-859-17

```

```

Query Match      83.2%; Score 15.8; DB 12; Length 33;
Best Local Similarity 89.5%; Pred. No. 67;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 ACAAGATACAGAACATGC 19
        ||||| ||||| |||||
Db      30 ACAAGCTACAAACAATGC 12

```

RESULT 3

```

; Sequence 9755, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agiate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9755
; LENGTH: 159
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9755

```

```

Query Match      83.2%; Score 15.8; DB 10; Length 159;
Best Local Similarity 89.5%; Pred. No. 84;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 ACAAGATACAGAACATGC 19
        ||||| ||||| |||||
Db      54 ACAAGATACAGACATGC 36

```

RESULT 4

```

; Sequence 8964, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agiate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 8964

```

```

; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-8964

```

```

Query Match      83.2%; Score 15.8; DB 10; Length 187;
Best Local Similarity 89.5%; Pred. No. 86;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 ACAAGATACAGAACATGC 19
        ||||| ||||| |||||
Db      65 ACAAGATACAGACATGC 47

```

RESULT 5

```

; Sequence 422, Application US/09925297
; Patent No. US20020081659A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: P105
; CURRENT APPLICATION NUMBER: US/09/925,297
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05983
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 928
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 422
; LENGTH: 366
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (264)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (327)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (348)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-925-297-422

```

```

Query Match      83.2%; Score 15.8; DB 10; Length 366;
Best Local Similarity 89.5%; Pred. No. 95;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY      1 ACAAGATACAGAACATGC 19
        ||||| ||||| |||||
Db      240 ACAAGATACAGAACATGC 258

```

RESULT 6

```

; Sequence 1999, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAMA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIRO
; APPLICANT: TATEISHI, NAOKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO

```

;; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;; FILE REFERENCE: 249-125
;; CURRENT APPLICATION NUMBER: US/09/738,626
;; CURRENT FILING DATE: 2000-12-18
;; PRIOR APPLICATION NUMBER: JP 99/377484
;; PRIOR FILING DATE: 1999-12-16
;; PRIOR APPLICATION NUMBER: JP 00/159162
;; PRIOR FILING DATE: 2000-04-07
;; PRIOR APPLICATION NUMBER: JP 00/280988
;; PRIOR FILING DATE: 2000-08-03
;; NUMBER OF SEQ ID NOS: 7059
;; SOFTWARE: PatentIn ver. 3.0
;; SEQ ID NO 1999
;; LENGTH: 375
;; TYPE: DNA
;; ORGANISM: Corynebacterium glutamicum
US-09-738-626-1999

Query Match 83.2%; Score 15.8; DB 9; Length 375;
Best Local Similarity 89.5%; Pred. No. 96;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGC 19
||||| |||||||
DB 216 ACAAGAGCAGACATGC 234

RESULT 7

US-10-011-859-1
;; Sequence 1, Application US/10011859
;; Patent No. US20020147328A1
;; GENERAL INFORMATION:
;; APPLICANT: Sheppard, Paul O.
;; TITLE OF INVENTION: CONNECTIVE TISSUE GROWTH FACTOR HOMOLOGS
;; FILE REFERENCE: 97-75
;; CURRENT APPLICATION NUMBER: US/10/011,859
;; CURRENT FILING DATE: 2001-11-05
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US/09/233,316
;; PRIOR FILING DATE: EARLIER FILING DATE: 1999-02-19
;; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: US 60/075,300
;; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-20
;; NUMBER OF SEQ ID NOS: 34
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 1
;; LENGTH: 1142
;; TYPE: DNA
;; ORGANISM: Homo sapiens
;; FEATURE:
;; NAME/KEY: CDS
;; LOCATION: (17)...(1078)
US-10-011-859-1

Query Match 83.2%; Score 15.8; DB 12; Length 1142;
Best Local Similarity 89.5%; Pred. No. 1.1e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGC 19
||||| |||||||
DB 584 ACAAGCTACAAACATGC 602

RESULT 8

US-09-864-864-293
;; Sequence 293, Application US/09864864
;; Patent No. US20020102679A1
;; GENERAL INFORMATION:
;; APPLICANT: Xu, Jiangchun
;; APPLICANT: Mitcham, Jennifer L.
;; APPLICANT: Harlocker, Susan L.
;; APPLICANT: Dillon, David C.
;; APPLICANT: Secrist, Heather
;; APPLICANT: Lodes, Michael J.

;; APPLICANT: Algate, Paul A.
;; APPLICANT: Fling, Steve P.
;; APPLICANT: Mannion, Jane
;; APPLICANT: Benson, Darin R.
;; APPLICANT: Carter, Darrick
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
;; FILE REFERENCE: 210121.523
;; CURRENT APPLICATION NUMBER: US/09/864,864
;; CURRENT FILING DATE: 2001-05-23
;; NUMBER OF SEQ ID NOS: 341
;; SOFTWARE: Corixa Invention Disclosure Database
;; SEQ ID NO 293
;; LENGTH: 6409
;; TYPE: DNA
;; ORGANISM: Homo sapiens
US-09-864-864-293

Query Match 83.2%; Score 15.8; DB 10; Length 6409;
Best Local Similarity 89.5%; Pred. No. 1.5e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGC 19
||||| |||||||
DB 4018 ACAAGATAAAGACATGC 4036

RESULT 9

US-09-820-905-3
;; Sequence 3, Application US/09820905
;; Patent No. US20020142938A1
;; GENERAL INFORMATION:
;; APPLICANT: Yan, Chunhua
;; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
;; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
;; FILE REFERENCE: C1001199
;; CURRENT APPLICATION NUMBER: US/09/820,905
;; CURRENT FILING DATE: 2001-03-30
;; NUMBER OF SEQ ID NOS: 5
;; SOFTWARE: FastSeq for Windows Version 4.0
;; SEQ ID NO 3
;; LENGTH: 203654
;; TYPE: DNA
;; ORGANISM: HUMAN
;; FEATURE:
;; NAME/KEY: misc_feature
;; LOCATION: (1)...(203654)
;; OTHER INFORMATION: n = A,T,C or G
US-09-820-905-3

Query Match 83.2%; Score 15.8; DB 10; Length 203654;
Best Local Similarity 89.5%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGC 19
||||| |||||||
DB 128516 AAAAGATACAGACATGC 128534

RESULT 10

US-09-770-444-72/C
;; Sequence 72, Application US/09770444
;; Patent No. US20020023280A1
;; GENERAL INFORMATION:
;; APPLICANT: Gorlach, Jorn
;; APPLICANT: An, Yong-Qiang
;; APPLICANT: Hamilton, Carol M.
;; APPLICANT: Price, Jennifer L.
;; APPLICANT: Raines, Tracy M.
;; APPLICANT: Yu, Yang
;; APPLICANT: Rameaka, Joshua G.
;; APPLICANT: Page, Amy

APPLICANT: Mathew, Abraham V.
APPLICANT: Ledford, Brooke L.
APPLICANT: Moessner, Jeffrey P.
APPLICANT: Haas, William David
APPLICANT: Garcia, Carlos A.
APPLICANT: Kicker, Maja
APPLICANT: Slader, Ted
APPLICANT: Davis, Keith R.
APPLICANT: Allen, Keith
APPLICANT: Hoffman, Neil
APPLICANT: Hurban, Patrick
TITLE OF INVENTION: Expressed Sequences of Arabidopsis
FILE REFERENCE: 2027 (PARA-016PRV)
CURRENT APPLICATION NUMBER: US/09/770,444
CURRENT FILING DATE: 2001-01-26
PRIORITY FILING DATE: 2000-01-27
NUMBER OF SEQ ID NOS: 999
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 72
LENGTH: 475
TYPE: DNA
ORGANISM: Arabidopsis thaliana
US-09-770-444-72

Query Match 81.1%; Score 15.4; DB 10; Length 475;
Best Local Similarity 94.1%; Pred. No. 1.5e+02;
Matches 16; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 2 CAAGATACAGACATG 18
DB 155 CAAGATACAGACATG 139

RESULT 11
US-09-731-231A-3
Sequence 3, Application US/09731231A
Patent No. US20020082189A1
GENERAL INFORMATION:
APPLICANT: GUEGLER, Karl et al
TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
FILE REFERENCE: CL001007
CURRENT APPLICATION NUMBER: US/09/731,231A
CURRENT FILING DATE: 2000-12-07
NUMBER OF SEQ ID NOS: 6
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 3
LENGTH: 326014
TYPE: DNA
ORGANISM: Human
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(326014)
OTHER INFORMATION: n = A,T,C or G
US-09-731-231A-3

Query Match 78.9%; Score 15; DB 10; Length 326014;
Best Local Similarity 100.0%; Pred. No. 4.9e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 4 AGATACAGACATG 18
DB 66273 AGATACAGACATG 66287

RESULT 12
US-10-046-935-1011
Sequence 1011, Application US/10046935
Patent No. US20020156011A1
GENERAL INFORMATION:

APPLICANT: Jiang, Yugu
APPLICANT: Harlocker, Susan L.
APPLICANT: Secrist, Heather
APPLICANT: Wang, Aijun
APPLICANT: Stolk, John A.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.527C1
CURRENT APPLICATION NUMBER: US/10/046,935
CURRENT FILING DATE: 2002-01-15
NUMBER OF SEQ ID NOS: 2239
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1011
LENGTH: 175
TYPE: DNA
ORGANISM: Homo sapiens
US-10-046-935-1011

Query Match 77.9%; Score 14.8; DB 9; Length 175;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGACATG 18
DB 92 ACAAGATACAGACATG 109

RESULT 13
US-09-878-178-1011
Sequence 1011, Application US/09878178
Patent No. US2002017552A1
GENERAL INFORMATION:
APPLICANT: Jiang, Yugu
APPLICANT: Harlocker, Susan L.
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.527
CURRENT APPLICATION NUMBER: US/09/878,178
CURRENT FILING DATE: 2001-06-08
NUMBER OF SEQ ID NOS: 2237
SOFTWARE: FASTSEQ for Windows Version 4.0
SEQ ID NO 1011
LENGTH: 175
TYPE: DNA
ORGANISM: Homo sapien
US-09-878-178-1011

Query Match 77.9%; Score 14.8; DB 9; Length 175;
Best Local Similarity 88.9%; Pred. No. 2.5e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATACAGACATG 18
DB 92 ACAAGATACAGACATG 109

RESULT 14
US-09-998-598-1959/c
Sequence 1959, Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database

SEQ ID NO 1959
LENGTH: 242
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-1959

Query Match 77.9%; Score 14.8; DB 10; Length 242;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 1 ACAAGATCAGACAATG 18
||||| ||||| |||
Db 238 ACAAGATCAGACAATG 221

RESULT 15
US-09-867-701-54
Sequence 54, Application US/09867701
Patent No. US20020132237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert.
APPLICANT: Hatlock, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
OF TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FASTSEQ FOR WINDOWS Version 4.0
SEQ ID NO 54
LENGTH: 272
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-54

Query Match 77.9%; Score 14.8; DB 10; Length 272;
Best Local Similarity 88.9%; Pred. No. 2.7e+02;
Matches 16; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 2 CAAGATCAGACAATGC 19
||||| ||||| |||
Db 84 CAAGATCAGACAATGC 101

Search completed: January 23, 2003, 20:35:52
Job time : 82.2857 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 19:25:32; Search time 165.029 seconds
(without alignments)
259.276 Million cell updates/sec

Title: US-09-803-736-1079_COPY_13662_13680
Perfect score: 19
Sequence: 1 acaagatacgaacatgac 19

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

Listing first 45 summaries

N.Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	17.4	91.6	1767	21 AAC49845	Arabidopsis thaliana
C 2	17.4	91.6	2476	21 AAC49848	Arabidopsis thaliana
C 3	17.4	91.6	2479	21 AAC37421	Arabidopsis thaliana
C 4	16.4	86.3	459	22 AAK77025	Human immune/haema
5	16.4	86.3	863	21 AAC54060	Arabidopsis thaliana
6	16.4	86.3	883	21 AAC53607	Arabidopsis thaliana
7	16.4	86.3	26040	22 AAH27887	Nucleotide sequenc
8	16.4	86.3	30600	20 AAC32023	Human METH1 relate
9	16.4	86.3	30601	22 AAC90080	Z69361 CDNA clone.

10	16.4	86.3	49999	20 AA23899	Human LOBO homolog
C 11	16	84.2	6659	21 AA23862	DNA encoding a GDN
C 12	15.8	83.2	33	20 AA271256	PCR primer ZC10422
C 13	15.8	83.2	33	24 AAK86311	Human connective t
C 14	15.8	83.2	141	22 AAK58235	Human immune/haema
C 15	15.8	83.2	159	24 ABL86777	Human ovarian cance
C 16	15.8	83.2	171	24 ABL87854	Human ORE3495 CDNA
C 17	15.8	83.2	187	24 ABL85966	Human ovarian cance
C 18	15.8	83.2	366	21 AAC99194	Human pancreatic c
C 19	15.8	83.2	367	22 AAK35358	Human colon cancer
C 20	15.8	83.2	375	22 AAK66964	C glutamine codin
C 21	15.8	83.2	475	22 AAK72268	Human immune/haema
C 22	15.8	83.2	543	24 ABL18131	Human OREX polyuac
C 23	15.8	83.2	599	21 AAA06933	Human growth facto
C 24	15.8	83.2	637	21 AAA06932	Human growth facto
C 25	15.8	83.2	705	23 AAS94211	DNA encoding novel
C 26	15.8	83.2	1076	23 AAS94210	DNA encoding novel
C 27	15.8	83.2	1142	20 AAZ11245	Human connective t
C 28	15.8	83.2	1142	24 AAK88298	DNA encoding human
C 29	15.8	83.2	1183	21 AAA06928	Human growth facto
C 30	15.8	83.2	1212	20 AAX76499	Human WISP-3 prote
C 31	15.8	83.2	1212	20 AAX76500	Human WISP-3 prote
C 32	15.8	83.2	1335	20 AAX76497	Human WISP-3 prote
C 33	15.8	83.2	1335	20 AAX76498	Human WISP-3 prote
C 34	15.8	83.2	1908	22 AAK94459	Human foetal CDNA,
C 35	15.8	83.2	2977	23 AAK82711	DNA encoding novel
C 36	15.8	83.2	4239	23 ABL18740	Drosophila melanog
C 37	15.8	83.2	4363	19 AAV23128	CDNA of protein w1
C 38	15.8	83.2	6409	24 ABR09756	Human ovarian tumo
C 39	15.8	83.2	9316	23 ABL27862	Drosophila melanog
C 40	15.8	83.2	18491	22 AAK65236	Human immune/haema
C 41	15.8	83.2	18495	22 AAK65235	Human immune/haema
C 42	15.8	83.2	28310	23 ABL16222	Drosophila melanog
C 43	15.8	83.2	34980	22 AAK68552	C glutamine codin
C 44	15.4	81.1	400	22 AAI89156	Human poly nucleoti
C 45	15.4	81.1	475	24 ABL93307	Arabidopsis thaliana

ALIGNMENTS

RESULT 1
AAC49845/c
AAC49845 standard; DNA; 1767 BP.

XX
AC AAC49845;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62651.
XX
KW Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.

OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 990S-0121825.
XX
PR 05-MAR-1999; 990S-0123180.
XX
PR 09-MAR-1999; 990S-0123548.
XX
PR 23-MAR-1999; 990S-0125788.
XX
PR 25-MAR-1999; 990S-0126264.
XX
PR 29-MAR-1999; 990S-0126785.
XX
PR 01-APR-1999; 990S-0127462.
XX
PR 06-APR-1999; 990S-0128234.
XX
PR 08-APR-1999; 990S-0128714.
XX
PR 16-APR-1999; 990S-0129845.

PR 19-APR-1999; 99US-0130077.
 PR 21-APR-1999; 99US-0130449.
 PR 23-APR-1999; 99US-0130510.
 PR 23-APR-1999; 99US-0130891.
 PR 28-APR-1999; 99US-0131449.
 PR 30-APR-1999; 99US-0132048.
 PR 30-APR-1999; 99US-0132407.
 PR 04-MAY-1999; 99US-0132484.
 PR 05-MAY-1999; 99US-0132485.
 PR 06-MAY-1999; 99US-0132486.
 PR 06-MAY-1999; 99US-0132487.
 PR 07-MAY-1999; 99US-0132863.
 PR 11-MAY-1999; 99US-0134256.
 PR 14-MAY-1999; 99US-0134218.
 PR 14-MAY-1999; 99US-0134219.
 PR 14-MAY-1999; 99US-0134221.
 PR 14-MAY-1999; 99US-0134370.
 PR 18-MAY-1999; 99US-0134378.
 PR 19-MAY-1999; 99US-0134441.
 PR 20-MAY-1999; 99US-0135124.
 PR 21-MAY-1999; 99US-0135353.
 PR 24-MAY-1999; 99US-0135629.
 PR 25-MAY-1999; 99US-0136021.
 PR 27-MAY-1999; 99US-0136392.
 PR 28-MAY-1999; 99US-0136782.
 PR 01-JUN-1999; 99US-0137222.
 PR 03-JUN-1999; 99US-0137528.
 PR 04-JUN-1999; 99US-0137502.
 PR 07-JUN-1999; 99US-0137724.
 PR 08-JUN-1999; 99US-0138094.
 PR 10-JUN-1999; 99US-0138540.
 PR 10-JUN-1999; 99US-0138847.
 PR 14-JUN-1999; 99US-0139119.
 PR 16-JUN-1999; 99US-0139452.
 PR 16-JUN-1999; 99US-0139453.
 PR 17-JUN-1999; 99US-0139492.
 PR 18-JUN-1999; 99US-0139454.
 PR 18-JUN-1999; 99US-0139455.
 PR 18-JUN-1999; 99US-0139456.
 PR 18-JUN-1999; 99US-0139457.
 PR 18-JUN-1999; 99US-0139458.
 PR 18-JUN-1999; 99US-0139459.
 PR 18-JUN-1999; 99US-0139460.
 PR 18-JUN-1999; 99US-0139461.
 PR 18-JUN-1999; 99US-0139462.
 PR 18-JUN-1999; 99US-0139463.
 PR 18-JUN-1999; 99US-0139750.
 PR 18-JUN-1999; 99US-0139763.
 PR 21-JUN-1999; 99US-0139817.
 PR 22-JUN-1999; 99US-0139899.
 PR 23-JUN-1999; 99US-0140353.
 PR 23-JUN-1999; 99US-0140354.
 PR 24-JUN-1999; 99US-0140685.
 PR 28-JUN-1999; 99US-0140823.
 PR 29-JUN-1999; 99US-0140991.
 PR 30-JUN-1999; 99US-0141287.
 PR 01-JUL-1999; 99US-0141842.
 PR 01-JUL-1999; 99US-0142154.
 PR 02-JUL-1999; 99US-0142055.
 PR 06-JUL-1999; 99US-0142380.
 PR 08-JUL-1999; 99US-0142803.
 PR 09-JUL-1999; 99US-0142920.
 PR 12-JUL-1999; 99US-0142977.
 PR 13-JUL-1999; 99US-0143542.
 PR 14-JUL-1999; 99US-0143624.
 PR 15-JUL-1999; 99US-0144005.
 PR 16-JUL-1999; 99US-0144085.
 PR 16-JUL-1999; 99US-0144096.
 PR 19-JUL-1999; 99US-0144325.
 PR 19-JUL-1999; 99US-0144331.
 PR 19-JUL-1999; 99US-0144332.
 PR 19-JUL-1999; 99US-0144333.
 PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
 PR 20-JUL-1999; 99US-0144352.
 PR 20-JUL-1999; 99US-0144632.
 PR 20-JUL-1999; 99US-0144884.
 PR 21-JUL-1999; 99US-0144814.
 PR 21-JUL-1999; 99US-0145086.
 PR 21-JUL-1999; 99US-0145088.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145087.
 PR 22-JUL-1999; 99US-0145089.
 PR 22-JUL-1999; 99US-0145192.
 PR 23-JUL-1999; 99US-0145145.
 PR 23-JUL-1999; 99US-0145218.
 PR 23-JUL-1999; 99US-0145224.
 PR 26-JUL-1999; 99US-0145276.
 PR 27-JUL-1999; 99US-0145812.
 PR 27-JUL-1999; 99US-0145918.
 PR 27-JUL-1999; 99US-0145919.
 PR 28-JUL-1999; 99US-0145951.
 PR 02-AUG-1999; 99US-0146386.
 PR 02-AUG-1999; 99US-0146388.
 PR 03-AUG-1999; 99US-0146389.
 PR 04-AUG-1999; 99US-0147038.
 PR 04-AUG-1999; 99US-0147204.
 PR 05-AUG-1999; 99US-0147302.
 PR 05-AUG-1999; 99US-0147192.
 PR 06-AUG-1999; 99US-0147260.
 PR 06-AUG-1999; 99US-0147303.
 PR 09-AUG-1999; 99US-0147416.
 PR 09-AUG-1999; 99US-0147493.
 PR 10-AUG-1999; 99US-0147935.
 PR 11-AUG-1999; 99US-0148177.
 PR 11-AUG-1999; 99US-0148319.
 PR 12-AUG-1999; 99US-0148341.
 PR 13-AUG-1999; 99US-0148565.
 PR 13-AUG-1999; 99US-0148684.
 PR 16-AUG-1999; 99US-0149368.
 PR 17-AUG-1999; 99US-0149175.
 PR 18-AUG-1999; 99US-0149426.
 PR 20-AUG-1999; 99US-0149722.
 PR 20-AUG-1999; 99US-0149723.
 PR 20-AUG-1999; 99US-0149929.
 PR 23-AUG-1999; 99US-0149902.
 PR 23-AUG-1999; 99US-0149930.
 PR 25-AUG-1999; 99US-0150566.
 PR 26-AUG-1999; 99US-0150884.
 PR 27-AUG-1999; 99US-0151065.
 PR 27-AUG-1999; 99US-0151066.
 PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151930.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 11-SEP-1999; 99US-0153758.
 PR 12-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154779.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157783.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159295.
 PR 14-OCT-1999; 99US-0159329.

```
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.6%; Score 17.4; DB 21; Length 1767;
Best Local Similarity 94.7%; Pred. No. 83;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGTACAGACATGC 19
Db 596 ACAAGAACGACATGC 578

RESULT 2
AAC49848/c
XX ID AAC49848 standard; DNA: 2476 BP.
AC AAC49848;
XX 18-OCT-2000 (first entry)
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 62661.
XX
XX Hybridisation assay; genetic mapping; gene expression control;
KW protein identification; signal transduction pathway;
KW metabolic pathway; promoter; termination sequence; ss.
XX
OS Arabidopsis thaliana.
XX
XX EPI033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-030139.
XX
XX 25-FEB-1999; 99US-012125.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130049.
XX 23-APR-1999; 99US-0130510.
XX 23-APR-1999; 99US-0130891.
XX 28-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134370.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0136029.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140623.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
```

PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145313.
PR 27-JUL-1999; 99US-0145319.
PR 28-JUL-1999; 99US-0145351.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 16-AUG-1999; 99US-0148684.
PR 17-AUG-1999; 99US-0149368.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 13-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.

PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.6%; Score 17.4; DB 21; Length 2476;
Best Local Similarity 94.7%; Pred. No. 85;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACGACATATGC 19
Db 685 ACAAGAAACGACATATGC 667

RESULT 3
AAC37421/C
ID AAC37421 standard; DNA: 2479 bp.
XX AC AAC37421;
XX DT 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 17316.
XX KM Hybridisation assay; genetic mapping; gene expression control;
XX KM protein identification; signal transduction pathway;
XX KM metabolic pathway; promoter; termination sequence; ss.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PE 25-FEB-2000; 2000BP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125768.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135253.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156559.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160778.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.

PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 91.6%; Score 17.4; DB 21; Length 2479;
Best Local Similarity 94.7%; Pred. No. 85;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGC 19
||||| |||||||
Db 687 ACAAGAAACAGACATGC 669

RESULT 4

AAK77025/c
ID AAK77025 standard; DNA; 459 bp.

XX AAK77025;

DT 07-NOV-2001 (first entry)

XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:31837.

KW Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;

KM cytosolic; gene therapy; vaccine; metastasis; ds.

XX Homo sapiens.

PN WC020157182-A2.

XX 09-AUG-2001.

PE 17-JAN-2001; 2001WO-US01354.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184654.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0188874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 11-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 14-JUL-2000; 2000US-0217496.

PR 26-JUL-2000; 2000US-0220963.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 18-AUG-2000; 2000US-0225759.

PR 22-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226868.

PR 23-AUG-2000; 2000US-0227182.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 06-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235834.

PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.

PR 29-SEP-2000; 2000US-0236367.

PR 29-SEP-2000; 2000US-0236368.

PR 29-SEP-2000; 2000US-0236369.

PR 29-SEP-2000; 2000US-0236370.

PR 02-OCT-2000; 2000US-0236802.

PR 02-OCT-2000; 2000US-0237037.

PR 02-OCT-2000; 2000US-0237039.

PR 02-OCT-2000; 2000US-0237040.

PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.

PR 20-OCT-2000; 2000US-0240960.

PR 20-OCT-2000; 2000US-0241221.

PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.

PR 20-OCT-2000; 2000US-0241787.

PR 20-OCT-2000; 2000US-0241808.

PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.

PR 01-NOV-2000; 2000US-0244617.

PR 08-NOV-2000; 2000US-0246474.

PR 08-NOV-2000; 2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.

PR 08-NOV-2000; 2000US-0246477.

PR 08-NOV-2000; 2000US-0246478.

PR 08-NOV-2000; 2000US-0246523.

PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.

PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.

PR 08-NOV-2000; 2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.

PR 08-NOV-2000; 2000US-0246610.

PR 08-NOV-2000; 2000US-0246611.

PR 08-NOV-2000; 2000US-0246613.

PR 17-NOV-2000; 2000US-0249207.


```
PR 18-JUN-1999; 99US-0139459;
PR 18-JUN-1999; 99US-0139460;
PR 18-JUN-1999; 99US-0139461;
PR 18-JUN-1999; 99US-0139462;
PR 18-JUN-1999; 99US-0139463;
PR 18-JUN-1999; 99US-0139750;
PR 18-JUN-1999; 99US-0139763;
PR 21-JUN-1999; 99US-0139817;
PR 22-JUN-1999; 99US-0139859;
PR 23-JUN-1999; 99US-0140353;
PR 23-JUN-1999; 99US-0140354;
PR 24-JUN-1999; 99US-0140695;
PR 28-JUN-1999; 99US-0140823;
PR 29-JUN-1999; 99US-0140991;
PR 30-JUN-1999; 99US-0141287;
PR 01-JUL-1999; 99US-0141842;
PR 01-JUL-1999; 99US-0142154;
PR 02-JUL-1999; 99US-0142055;
PR 06-JUL-1999; 99US-0142390;
PR 08-JUL-1999; 99US-0142803;
PR 09-JUL-1999; 99US-0142920;
PR 12-JUL-1999; 99US-0142977;
PR 13-JUL-1999; 99US-0143542;
PR 14-JUL-1999; 99US-0143524;
PR 15-JUL-1999; 99US-0144003;
PR 16-JUL-1999; 99US-0144085;
PR 16-JUL-1999; 99US-0144086;
PR 19-JUL-1999; 99US-0144325;
PR 19-JUL-1999; 99US-0144331;
PR 19-JUL-1999; 99US-0144332;
PR 19-JUL-1999; 99US-0144333;
PR 19-JUL-1999; 99US-0144334;
PR 19-JUL-1999; 99US-0144335;
PR 20-JUL-1999; 99US-0144352;
PR 20-JUL-1999; 99US-0144632;
PR 20-JUL-1999; 99US-0144684;
PR 21-JUL-1999; 99US-0144814;
PR 21-JUL-1999; 99US-0145086;
PR 21-JUL-1999; 99US-0145088;
PR 22-JUL-1999; 99US-0145085;
PR 22-JUL-1999; 99US-0145087;
PR 22-JUL-1999; 99US-0145089;
PR 22-JUL-1999; 99US-0145192;
PR 23-JUL-1999; 99US-0145145;
PR 23-JUL-1999; 99US-0145218;
PR 23-JUL-1999; 99US-0145224;
PR 26-JUL-1999; 99US-0145376;
PR 27-JUL-1999; 99US-0145913;
PR 27-JUL-1999; 99US-0145918;
PR 27-JUL-1999; 99US-0145919;
PR 28-JUL-1999; 99US-0145951;
PR 02-AUG-1999; 99US-0146386;
PR 02-AUG-1999; 99US-0146388;
PR 02-AUG-1999; 99US-0146389;
PR 03-AUG-1999; 99US-0147038;
PR 04-AUG-1999; 99US-0147204;
PR 04-AUG-1999; 99US-0147302;
PR 05-AUG-1999; 99US-0147192;
PR 05-AUG-1999; 99US-0147260;
PR 06-AUG-1999; 99US-0147303;
PR 06-AUG-1999; 99US-0147416;
PR 09-AUG-1999; 99US-0147493;
PR 09-AUG-1999; 99US-0147935;
PR 10-AUG-1999; 99US-0148171;
PR 11-AUG-1999; 99US-0148319;
PR 12-AUG-1999; 99US-0148341;
PR 13-AUG-1999; 99US-0148665;
PR 13-AUG-1999; 99US-0148684;
PR 16-AUG-1999; 99US-0149368;
PR 17-AUG-1999; 99US-0149175;
PR 18-AUG-1999; 99US-0149426;
PR 20-AUG-1999; 99US-0149722;
PR 20-AUG-1999; 99US-0149723;

PR 20-AUG-1999; 99US-0149929;
PR 23-AUG-1999; 99US-0149902;
PR 23-AUG-1999; 99US-0149930;
PR 25-AUG-1999; 99US-0150566;
PR 26-AUG-1999; 99US-0150884;
PR 27-AUG-1999; 99US-0151065;
PR 27-AUG-1999; 99US-0151066;
PR 27-AUG-1999; 99US-0151080;
PR 30-AUG-1999; 99US-0151303;
PR 31-AUG-1999; 99US-0151308;
PR 01-SEP-1999; 99US-0151930;
PR 07-SEP-1999; 99US-0152363;
PR 10-SEP-1999; 99US-0153070;
PR 13-SEP-1999; 99US-0153758;
PR 15-SEP-1999; 99US-0154018;
PR 16-SEP-1999; 99US-0154039;
PR 20-SEP-1999; 99US-0154779;
PR 22-SEP-1999; 99US-0155139;
PR 23-SEP-1999; 99US-0155486;
PR 24-SEP-1999; 99US-0155659;
PR 28-SEP-1999; 99US-0156458;
PR 29-SEP-1999; 99US-0156556;
PR 04-OCT-1999; 99US-0157117;
PR 05-OCT-1999; 99US-0157753;
PR 06-OCT-1999; 99US-0157865;
PR 07-OCT-1999; 99US-0158029;
PR 08-OCT-1999; 99US-0158232;
PR 12-OCT-1999; 99US-0158369;
PR 13-OCT-1999; 99US-0159293;
PR 13-OCT-1999; 99US-0159294;
PR 13-OCT-1999; 99US-0159295;
PR 14-OCT-1999; 99US-0159329;
PR 14-OCT-1999; 99US-0159330;
PR 14-OCT-1999; 99US-0159331;
PR 14-OCT-1999; 99US-0159637;
PR 14-OCT-1999; 99US-0159638;
PR 18-OCT-1999; 99US-0159584;
PR 21-OCT-1999; 99US-0160741;
PR 21-OCT-1999; 99US-0160767;
PR 21-OCT-1999; 99US-0160768;
PR 21-OCT-1999; 99US-0160770;
PR 21-OCT-1999; 99US-0160814;
PR 21-OCT-1999; 99US-0160815;
PR 22-OCT-1999; 99US-0160980;
PR 22-OCT-1999; 99US-0160981;
PR 22-OCT-1999; 99US-0160989;
PR 25-OCT-1999; 99US-0161404;
PR 25-OCT-1999; 99US-0161405;
PR 25-OCT-1999; 99US-0161406;
PR 26-OCT-1999; 99US-0161359;
PR 26-OCT-1999; 99US-0161360;
PR 26-OCT-1999; 99US-0161361;
PR 28-OCT-1999; 99US-0161920;
PR 28-OCT-1999; 99US-0161992;
PR 28-OCT-1999; 99US-0161993;
PR 29-OCT-1999; 99US-0162142;

QY 1 ACAGATTCAGAACATG 18
      |||||
Db 782 ACAGATTCAGAACATTG 799

Query Match 86.3%; Score 16.4; DB 21; Length 863;
Best Local Similarity 94.4%; Pred. No. 2.3e+02;
Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

RESULT 6
AACG3607
ID AACG3607 standard; DNA; 863 BP.
XX
AC AACG3607;
XX
DT 18-OCT-2000 (first entry)
```


PR 27-AUG-1999; 99US-0151080.
 PR 30-AUG-1999; 99US-0151303.
 PR 31-AUG-1999; 99US-0151438.
 PR 01-SEP-1999; 99US-0151920.
 PR 07-SEP-1999; 99US-0152363.
 PR 10-SEP-1999; 99US-0153070.
 PR 13-SEP-1999; 99US-0153758.
 PR 15-SEP-1999; 99US-0154018.
 PR 16-SEP-1999; 99US-0154039.
 PR 20-SEP-1999; 99US-0154739.
 PR 22-SEP-1999; 99US-0155139.
 PR 23-SEP-1999; 99US-0155486.
 PR 24-SEP-1999; 99US-0155659.
 PR 28-SEP-1999; 99US-0156458.
 PR 29-SEP-1999; 99US-0156596.
 PR 04-OCT-1999; 99US-0157117.
 PR 05-OCT-1999; 99US-0157753.
 PR 06-OCT-1999; 99US-0157865.
 PR 07-OCT-1999; 99US-0158029.
 PR 08-OCT-1999; 99US-0158232.
 PR 12-OCT-1999; 99US-0158369.
 PR 13-OCT-1999; 99US-0159293.
 PR 13-OCT-1999; 99US-0159294.
 PR 13-OCT-1999; 99US-0159329.
 PR 14-OCT-1999; 99US-0159330.
 PR 14-OCT-1999; 99US-0159331.
 PR 14-OCT-1999; 99US-0159637.
 PR 14-OCT-1999; 99US-0159638.
 PR 18-OCT-1999; 99US-0159584.
 PR 21-OCT-1999; 99US-0160741.
 PR 21-OCT-1999; 99US-0160767.
 PR 21-OCT-1999; 99US-0160768.
 PR 21-OCT-1999; 99US-0160770.
 PR 21-OCT-1999; 99US-0160814.
 PR 21-OCT-1999; 99US-0160815.
 PR 22-OCT-1999; 99US-0160980.
 PR 22-OCT-1999; 99US-0160981.
 PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161992.
 PR 28-OCT-1999; 99US-0161993.
 PR 29-OCT-1999; 99US-0162142.

Query Match 86.3%; Score 16.4; DB 21; Length 883;
 Best Local Similarity 94.4%; Pred. No. 2.3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACAAATG 18
 DB 802 ACAAGATACAGACAAATG 819

RESULT 7

AAH27887

ID AAH27887 standard; DNA; 26040 BP.
 AC AAH27887;
 XX
 DT 05-SEP-2001 (first entry)
 DE Nucleotide sequence of FRA16D oxidoreductase (FOR) gene fragment.

XX Cancer associated protein; FOR gene; FRA16D; fragile site; aphidicolin;
 KW chromosomal rearrangement; cancer; splice variant; DNA instability;
 KW FRA16D oxidoreductase; neoplasia; ss.
 XX

OS Homo sapiens.
 XX
 XX WO200144466-A1.
 XX
 XX 21-JUN-2001.
 PD
 PF 15-DEC-2000; 2000WO-AU01539.
 XX
 XX 16-DEC-1999; 99AU-0004711.
 PR 19-APR-2000; 2000AU-0007025.
 XX
 XX (WOMEN-) WOMEN'S & CHILDREN'S HOSPITAL.
 PA
 PI Richards R, Ried K, Finnis M, Hobson L, Mangelsdorf M, Dayan S;
 PI Nancarrow J, Woollatt E, Baker E;
 XX
 DR WPI; 2001-398151/42.
 XX
 PR Novel isolated 16q23.2 nucleic acid molecule, FRA16D oxidoreductase
 PR (FOR) gene associated with FRA16D site, useful for early diagnosis and
 PR assessment of risk of cancers associated with the FRA16D region
 XX
 PS Disclosure; Fig 12; 150pp; English.

CC The present sequence represents a fragment of the human FRA16D
 CC oxidoreductase (FOR) gene, encompassing the FRA16D fragile site. The FOR
 CC gene encodes a cancer associated protein. The FRA16D fragile site is
 CC induced by aphidicolin, which is located within the FOR gene. The fragile
 CC site is the location of breakpoints of a variety of chromosomal
 CC rearrangements, and other mutations associated with cancers. The FOR
 CC protein is expressed as a number of splice variants. FOR gene
 CC polynucleotide fragments are capable of acting as specific primers or
 CC probes for detecting cancer associated variations of DNA sequence such
 CC as a point mutation or small DNA rearrangement associated with the
 CC tumour, a breakpoint of one or more chromosomal rearrangements associated
 CC with the tumour and a pause site within the FRA16 gene. FOR nucleic acid
 CC molecules are useful as markers to identify relationship between the
 CC fragile site (FRA16D) and the DNA instability in neoplasia which allows
 CC better diagnosis of cancers associated with the region.
 CC
 XX
 S0 Sequence 26040 BP; 7255 A; 5270 C; 5690 G; 7825 T; 0 other;

Query Match 86.3%; Score 16.4; DB 22; Length 26040;
 Best Local Similarity 94.4%; Pred. No. 2.9e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2 CAAGATACAGACAAATGC 19
 DB 22863 CAAGATACAGACAAATGC 22880

RESULT 8

AAZ32023

AAZ32023

10-JAN-2000 (first entry)

Human MERTH related EST Z69361.

XX Human; MERTH1; MERTH2; anti-angiogenic; metalloprotease thrombospondin;
 KW cancer; diagnosis; hyperproliferative disorder; autoimmune disease;
 KW angiogenesis inhibitor; abnormal wound healing; inflammation;
 KW rheumatoid arthritis; psoriasis; endometrial bleeding disorder;
 KW diabetic retinopathy; macula degeneration; haemangioma; detection;
 KW arterial-venous malformation; immune deficiency; ss.

XX Homo sapiens.

XX
 XX WO9337660-A1.
 XX 29-JUL-1999.

XX 22-JAN-1999; 99WO-US01313.
 XX
 XX 23-JAN-1998; 98US-0072298.
 PR 28-AUG-1998; 98US-0098539.
 XX
 XX (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM;
 DR WPI: 1999-590684/50.
 XX
 PT New isolated metalloprotease thrombospondin polypeptides, useful for
 PT treating hyperproliferative disorders, cancers or autoimmune disorders
 PT
 PS Disclosure: Page 327-347; 457pp; English.
 XX
 XX AA32000 and AA32001 encode, and AAY49501 and AAY49502 represent, human
 CC metalloprotease thrombospondin (METH) proteins METH1 and METH2
 CC respectively. METH1 and METH2 have been found to be potent inhibitors of
 CC angiogenesis both in vitro and in vivo. They can be used for treating
 CC cancer and other disorders related to angiogenesis including abnormal
 CC wound healing, inflammation, rheumatoid arthritis, psoriasis,
 CC endometrial bleeding disorders, diabetic retinopathy, some forms of
 CC macula degeneration, haemangiomas, and arterial-venous malformations.
 CC They may be useful in treating deficiencies or disorders of the immune
 CC system, by activating or inhibiting the proliferation, differentiation,
 CC or mobilisation (chemotaxis) of immune cells. The etiology of these
 CC immune deficiencies or disorders may be genetic, somatic, such as
 CC cancer or some autoimmune disorders, acquired (e.g. by chemotherapy or
 CC toxins), or infectious. They can also be used to treat inflammatory or
 CC conditions, both chronic and acute conditions. The products can also be
 CC used for detection and diagnosis. AA32002 to AA32080, and AAY49503 to
 CC AAY49511 represent sequences given in the exemplification of the present
 CC invention.
 XX
 XX Sequence 30600 BP; 10390 A; 5313 C; 5429 G; 9467 T; 1 other:
 SQ
 Query Match 86.3%; Score 16.4; DB 20; Length 30600;
 Best Local Similarity 94.4%; Pred. No. 3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ACAAGATACAGACACATG 18
 DB 1493 ACAAGATACAGACACATG 1510
 RESULT 9
 AAC90080
 ID AAC90080 standard; DNA; 30601 BP.
 XX
 XX AAC90080;
 AC
 XX
 DT 19-MAR-2001 (first entry)
 DE 269361 cDNA clone.
 XX
 XX METH: metalloprotease; thrombospondin; angiogenesis inhibition;
 KW cancer therapy; benign tumour; ocular angiogenic disease;
 KW rheumatoid arthritis; psoriasis; wound healing; endometriosis;
 KW vasculogenesis; granulation; hypertrophic scar; nonunion fracture;
 KW scleroderma; tracheoma; vascular adhesion; myocardial angiogenesis;
 KW coronary collateral; cerebral collateral; arteriovenous malformation;
 KW ischaemic limb angiogenesis; Osler-Weber syndrome; wound granulation;
 KW plaque neovascularisation; telangiectasia; haemophilic joint; EST;
 KW angiodioma; fibromuscular dysplasia; expressed sequence tag;
 KW Crohn's disease; atherosclerosis; birth control; ss.
 XX
 XX Unidentified.
 OS
 XX

PN WO200071577-A1.
 PD 30-NOV-2000.
 XX
 XX 25-MAY-2000; 2000WO-US14462.
 PF
 XX 25-MAY-1999; 99US-0318208.
 PR 20-JUL-1999; 99US-0144882.
 PR 10-AUG-1999; 99US-0147823.
 PR 13-AUG-1999; 99US-0373658.
 PR 22-DEC-1999; 99US-0171503.
 PR 22-FEB-2000; 2000US-0183792.
 XX
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA (SMIK-) SMITHKLINE BEECHAM CORP.
 PA (BETH-) BETH ISRAEL DEACONESS MEDICAL CENT.
 PA (IRUE/) IRUELA-ARISPE L.
 PA (HAST/) HASTINGS G A.
 PA (RUBE/) RUBEN S M.
 PA (JONAK/) JONAK Z L.
 PA (TRULL/) TRULLI S H.
 PA (FORN/) FORNWALD J A.
 PA (TERR/) TERRRETT J A.
 XX
 PI IrueLA-Arispe L, Hastings GA, Ruben SM, Jonak ZL, Trulli SH;
 PI Fornwald JA, Terrrett JA;
 DR WPI: 2001-025136/03.
 XX
 XX METH1 and METH2 polynucleotides and encoded polypeptides, used to
 PT inhibit angiogenesis in the treatment of disorders such as cancer,
 PT rheumatoid arthritis and psoriasis -
 PT
 PS Claim 7; Pages 628-647; 768pp; English.
 XX
 XX The present invention relates to human METH1 and METH2, (ME for
 CC metalloprotease and TH for thrombospondin; see AAB50002 and AAB50003).
 CC The present sequence is an expressed sequence tag (EST) for METH. METH
 CC can be used for inhibiting angiogenesis in an individual, and for
 CC treating cancer, benign tumours, an ocular angiogenic disease,
 CC rheumatoid arthritis, psoriasis, delayed wound healing, endometriosis,
 CC vasculogenesis, granulations, hypertrophic scars, nonunion fractures,
 CC scleroderma, tracheoma, vascular adhesion, myocardial angiogenesis,
 CC coronary collaterals, cerebral collaterals, arteriovenous malformations,
 CC ischaemic limb angiogenesis, Osler-Weber syndrome, plaque
 CC neovascularisation, telangiectasia, haemophilic joints, angiodioma,
 CC fibromuscular dysplasia, wound granulation, Crohn's disease or
 CC atherosclerosis. METH can also be used in birth control. METH can also
 CC be used in diagnostic methods for the prognosis of cancer.
 XX
 XX Sequence 30601 BP; 10390 A; 5314 C; 5430 G; 9467 T; 0 other;
 SQ
 Query Match 86.3%; Score 16.4; DB 22; Length 30601;
 Best Local Similarity 94.4%; Pred. No. 3e+02;
 Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 ACAAGATACAGACACATG 18
 DB 1493 ACAAGATACAGACACATG 1510
 RESULT 10
 AA223899
 ID AA223899 standard; DNA; 49999 BP.
 XX
 XX AA223899;
 AC
 XX
 DT 25-JUN-2000 (first entry)
 DE Human LOBO homologue genomic DNA.
 XX
 XX LOBO: long bones; bone development; bone extension; skull; osteopathic;
 KW diagnostic; pharmaceutical; gene therapy; transgenic animal; disease;
 KW

```

KM spondyloepiphyseal dysplasia; achondroplasia; human; ds.
XX
XX Homo sapiens.
XX
XX W0950284-A2.
XX
XX 07-OCT-1999.
XX
XX 26-MAR-1999; 99MO-EP02055.
XX
XX 27-MAR-1999; 98DE-1013799.
XX
XX (ROSE/) ROSENTHAL A.
XX
XX Rosenthal A, Rump A, Hess J, Aigner T, Wirth T;
XX
XX WPI: 1999-601320/51.
XX
XX Nucleic acids encoding proteins which influence bone development,
XX
XX useful for treating and studying bone disorders -
XX
XX Example 3; Page 215-243; 391pp; German.
XX
XX This invention describes novel nucleic acids (I; designated LOBO (long
XX
XX bones)) encoding proteins influencing bone development in mammals. The
XX
XX proteins of the invention reduce and/or inactivate bone extension (i.e.
XX
XX development), with exception of the skull and have osteopathic activity.
XX
XX The nucleic acid molecules, proteins and antibodies can be used in
XX
XX diagnostic or pharmaceutical compounds e.g. for gene therapy. The methods
XX
XX and nucleic acid molecules, etc. are useful for production of transgenic
XX
XX animals, especially a transgenic mouse for the study of diseases
XX
XX associated with bone development, e.g. spondyloepiphyseal dysplasia and
XX
XX achondroplasia. This sequence encodes a human LOBO protein described
XX
XX in the method of the invention.
XX
XX Sequence 49999 BP; 13475 A; 10783 C; 11425 G; 14316 T; 0 other;
XX
XX
XX Query Match 86.3%; Score 16.4; DB 20; Length 49999;
XX
XX Best Local Similarity 94.4%; Pred. No. 3.1e+02;
XX
XX Matches 17; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
XX
XX 2 CAGATACGACGACATTC 19
XX
XX ||||||||||||||||
XX
XX Db 43070 CAGATACGACGACATTC 43087
XX
XX
XX RESULT 11
XX
XX ID AA36862/C
XX
XX AA36862 standard; DNA; 6659 BP.
XX
XX
XX AC AA36862;
XX
XX
XX DT 13-MAR-2000 (first entry)
XX
XX
XX DE DNA encoding a GON-1 protein of Caenorhabditis elegans.
XX
XX
XX KW GON-1; metalloproteinase; cell migration; modulator;
XX
XX metalloproteinase domain; thrombospondin domain; abnormal cell migration;
XX
XX organ shaping; sterility; cancer metastasis; ss.
XX
XX
XX OS Caenorhabditis elegans.
XX
XX
XX FH Key
XX
XX FT 1.6453
XX
XX CDS /tag= a
XX
XX /product= "GON-1"
XX
XX
XX W09961656-A2.
XX
XX
XX 02-DEC-1999.
XX
XX
XX 28-MAY-1999; 99MO-US11918.
XX
XX

```

```

PR 29-MAY-1998; 98US-0087170.
XX
XX 13-APR-1999; 99US-0129023.
XX
XX (WISC ) WISCONSIN ALUMNI RES FOUND.
XX
XX
XX Kimble JE, Belloch RH;
XX
XX WPI: 2000-072633/06.
XX
XX P-PSDB; AAY53898.
XX
XX
XX Identifying modulators of proteins containing metalloprotease and
XX
XX thrombospondin domains, potentially useful for controlling cell
XX
XX migration and organ shaping -
XX
XX Disclosure; Page 39-51; 60pp; English.
XX
XX
XX The present sequence encodes a GON-1 protein of Caenorhabditis elegans.
XX
XX GON-1 is a secreted metalloproteinase that lacks a transmembrane domain
XX
XX and possesses a predicted metalloprotease domain between amino acids
XX
XX 269-456. In C. elegans hermaphrodites, GON-1 is required for migration of
XX
XX two distal tip cells to produce elongated tubes, whereas in males, GON-1
XX
XX is required for migration of a single linker cell to produce a single
XX
XX elongated tube. The protein is used in the method of the invention. The
XX
XX specification describes a method for identifying a modulator of a
XX
XX protein that contains a metalloproteinase domain and a thrombospondin
XX
XX domain. The method comprises treating a target organism, having a test
XX
XX compound, and determining any change in migration or shape of the cell
XX
XX attributable to the test compound. The compounds identified are
XX
XX potential therapeutic modulators of abnormal cell migration and organ
XX
XX shaping, e.g. for rendering animals (specifically nematodes) sterile
XX
XX and for inhibiting cancer metastases.
XX
XX
XX Sequence 6659 BP; 1928 A; 1337 C; 1547 G; 1847 T; 0 other;
XX
XX
XX Query Match 84.2%; Score 16; DB 21; Length 6659;
XX
XX Best Local Similarity 100.0%; Pred. No. 4.1e+02;
XX
XX Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
XX 1 ACAAGATCAGACGACAA 16
XX
XX ||||||||||||||||
XX
XX Db 3723 ACAAGATCAGACGACAA 3708
XX
XX
XX RESULT 12
XX
XX ID AA211258/C
XX
XX AA211258 standard; DNA; 33 BP.
XX
XX
XX AC AA211258;
XX
XX
XX DT 15-NOV-1999 (first entry)
XX
XX
XX DE PCR primer ZC16422 for human ZCTGF4, coding sequence.
XX
XX
XX KW Connective tissue growth factor; ZCTGF4; chromosome 6q abnormality;
XX
XX diagnosis; therapy; bone marrow fibrosis; haematopoietic cell production;
XX
XX haematopoietic cell differentiation; scar tissue formation; scleroderma;
XX
XX cutaneous lupus erythematosus; dermatosis; end-stage kidney failure;
XX
XX human; PCR primer; ss.
XX
XX
XX OS Synthetic.
XX
XX
XX OS Homo sapiens.
XX
XX
XX W09942583-A1.
XX
XX
XX 26-AUG-1999.
XX
XX
XX 19-FEB-1999; 99MO-US03585.
XX
XX
XX 20-FEB-1999; 98US-0027405.
XX
XX
XX (ZYMO ) ZYMOGENETICS INC.
XX
XX

```

PI Jaspers SR, Sheppard PO;
 XX WPI; 1999-540308/45.
 DR
 XX
 PT New isolated polynucleotides encoding connective tissue growth
 factor homologue polypeptides, used diagnostically
 PT
 XX
 PS Example 4; Page 109; 122pp; English.
 CC This sequence represents a PCR primer for DNA encoding the human
 CC connective tissue growth factor (ZCTGF4) of the invention. The ZCTGF4
 CC coding sequence may be used to produce CTFP homologue polypeptides
 CC according to standard recombinant DNA methodologies. The ZCTGF4 DNAs may
 CC also be used diagnostically as probes to detect the presence of similar
 CC sequences in biological samples, and to identify abnormalities or
 CC mutations within those sequences. The sequences can be used to detect a
 CC chromosome 6q abnormality. Antagonists and antibodies of ZCTGF4 can be
 CC used to treat and diagnose bone marrow fibrosis, modulating production or
 CC differentiation of haematopoietic cells, prevention of scar tissue
 CC formation, cutaneous lupus erythematosus, scleroderma, dermatosis, and
 CC end-stage kidney failure.
 CC
 SO Sequence 33 BP; 6 A; 4 C; 8 G; 15 T; 0 other;
 Query Match 83.2%; Score 15.8; DB 20; Length 33;
 Best Local Similarity 89.5%; Pred. No. 3.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ACAGATACAGACAAATGC 19
 ||||| ||||| ||||| |||||
 Db 30 ACAGCTACAAAACAAATGC 12
 RESULT 13
 ID ABR88311/C
 XX ABR88311 standard; DNA; 33 BP.
 AC ABR88311;
 XX
 XX 07-OCT-2002 (first entry)
 DE Human connective tissue growth factor ZCTGF4 PCR primer ZC 16422.
 XX
 KM Human; connective tissue growth factor; CTFP; ZCTGF4; aneuploidy;
 KM chromosomal disorder; gene copy number change; insertion; deletion;
 KM restriction site change; restriction site rearrangement; testis;
 KM pathological condition; bone marrow fibrosis; scleroderma; trachea;
 KM scar tissue formation; cutaneous lupus erythematosus; dermatosis;
 KM end-stage kidney failure; gene therapy; PCR; primer; ss.
 XX
 OS Homo sapiens.
 XX
 PN US6395890-B1.
 XX
 PD 28-MAY-2002.
 XX
 PE 19-FEB-1999; 9905-0253316.
 XX
 PR 20-FEB-1998; 9805-075300P.
 XX
 PA (ZYMO) ZYMOGENETICS INC.
 XX
 PI Sheppard PO, Jaspers SR, Gao Z;
 XX
 DR WPI; 2002-556106/59.
 XX
 PT New polynucleotide acid encoding connective tissue growth factor
 PT homologue polypeptide, ZCTGF4, useful for producing an antagonist for
 PT treating/preventing pathological disorders e.g. scleroderma, and
 PT dermatosis
 XX
 PS Example 4; Column 59; 40pp; English.
 XX

CC The present invention relates to a new polynucleotide molecule encoding
 CC a connective tissue growth factor homologue polypeptide (ZCTGF4). The
 CC invention is useful for diagnosing chromosomal disorders (e.g.
 CC aneuploidy, gene copy number changes, insertions, deletions,
 CC restriction site changes and rearrangements) associated with abnormal
 CC expression of ZCTGF4 protein. The invention is also useful for analysing
 CC chromosomal DNA which is useful for correlating disease with
 CC abnormalities localised to chromosome 6. ZCTGF4 is useful for regulating
 CC the growth and/or differentiation of ZCTGF4 responsive cells, in treating
 CC disorders associated with upregulated growth in ZCTGF4-responsive
 CC tissues, and producing an antagonist to treat or prevent development of
 CC pathological conditions in tissues, such as, testis, trachea, bone
 CC marrow fibrosis, prevention of scar tissue formation, cutaneous lupus
 CC erythematosus, scleroderma, dermatosis and end-stage kidney failure.
 CC The invention is also useful in gene therapy. The present nucleic acid
 CC sequence represents a human connective tissue growth factor ZCTGF4
 CC PCR primer that was used to illustrate the invention.
 CC
 SO Sequence 33 BP; 6 A; 4 C; 8 G; 15 T; 0 other;
 Query Match 83.2%; Score 15.8; DB 24; Length 33;
 Best Local Similarity 89.5%; Pred. No. 3.6e+02;
 Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 ACAGATACAGACAAATGC 19
 ||||| ||||| ||||| |||||
 Db 30 ACAGCTACAAAACAAATGC 12
 RESULT 14
 ID AAK58235/C
 XX AAK58235 standard; cDNA; 141 BP.
 AC AAK58235;
 XX
 XX 06-NOV-2001 (first entry)
 DE Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:3295.
 KM Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
 KM cytostatic; gene therapy; vaccine; metastasis; ss.
 XX
 OS Homo sapiens.
 XX
 PN W0200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PE 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 15-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR

PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234224.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236337.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.

PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0251989.
PR 06-DEC-2000; 2000US-0251719.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251859.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

(HUMA) HUMAN GENOME SCI INC.
Rosen CA, Barash SC, Ruben SM.
WPI: 2001-483426/52.
P-PsDB: AAK65454.

Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
useful for preventing, diagnosing and/or treating cancers and
metastasis -
XX
PS Claim 1; SEQ ID NO 3295; 3071bp + Sequence Listing; English.
XX
AAK4951 to AAK64702 encode the human immune/hematopoietic antigen (I)
amino acid sequences given in AAK62170 to AAK61921. (I) have cytostatic
activity, and can be used in gene therapy and vaccine production. (I)
proteins and polynucleotides may be used in the prevention, diagnosis and
treatment of diseases associated with inappropriate (I) expression. For
example, they may be used to treat disorders associated with decreased
expression by rectifying mutations or deletions in a patient's genome
that affect the activity of (I) by expressing inactive proteins or to
supplement the patient's own production of (I). Additionally, (I)
polynucleotides may be used to produce the secreted (I), by inserting
the nucleic acids into a host cell and culturing the cell to express the
protein. (I) proteins and polynucleotides may be used to prevent,
diagnose and treat immune/hematopoietic-related diseases, especially
cancers and cancer metastases of hematopoietic-derived cells. AAK64703
to AAK67694 represent human immune/hematopoietic antigen genomic
sequences from the present invention. AAK51942 to AAK54950 and AAK62169
represent sequences used in the exemplification of the present invention.

SO Sequence 141 BP; 40 A; 27 C; 15 G; 59 T; 0 other;

Query Match

Best Local Similarity 83.2%; Score 15.8; DB 22; Length 141;
Pred. No. 3.9e+02;

Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 ACAAGATACAGACATGTC 19
||||| ||||||| |||
DB 55 ACAAGTTACAGACATGTC 37

QY 1 ACAAGATACAGACATGTC 19
||||| ||||||| |||
DB 54 ACAAGTTACAGACATGTC 36
Search completed: January 23, 2003, 21:27:22
Job time : 177.029 secs

RESULT 15

ABL86777/c

ID ABL86777 standard; cDNA: 159 BP.

XX ABL86777;

XX 17-MAY-2002 (first entry)

XX Human ovarian cancer related cDNA clone SEQ ID NO:9755.

XX Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.

XX Homo sapiens.

XX WO200192581-A2.

XX 06-DEC-2001.

XX 29-MAY-2001; 2001WO-US17756.

XX 26-MAY-2000; 2000US-207484P.

XX (CORI-) CORIXA CORP.

XX Algate PA, Harlocker SL, Jones R;

XX WPI: 2002-122075/16.

XX Composition for therapy and diagnosis of ovarian cancer comprising

PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding

PT polypeptide, antibody specific to polypeptide or T cell expressing

PS Claim 1; SEQ ID 9755; 489pp; English.

XX The present invention describes a composition (I) comprising: carriers

CC and immunostimulants; and a polypeptide (II) of a ovarian tumour

CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence

CC (SI) from the 10912 nucleotide sequences as given in ABL77023 to

CC ABL87934, (III) encoding (II) having a sequence (S2), a T cell

CC population of (II), or antigen presenting cells that express (II).

CC (I) has cytostatic activity. An oligonucleotide (IV) that hybridises to

CC sample preferably serum or ovarian tissue. The method comprises

CC contacting a biological sample from a patient with (IV), detecting the

CC amount of polynucleotide hybridising to (IV) and comparing the amount to

CC a predetermined cutoff value and thereby detecting ovarian cancer in the

CC patient, where the amount of polynucleotide hybridising to (IV) is

CC detected preferably by polymerase chain reaction (PCR). (I) comprising

CC (III) and/or (II) is useful for stimulating and/or expanding T cells

CC specific for an ovarian tumour protein comprising contacting T cells

CC with (III) or (II). (II) is useful in design and preparation of

CC ribozyme molecules for inhibiting expression of the tumour polypeptides

CC and proteins in tumour cells; and to isolate a full length gene from a

CC suitable library e.g., a tumour cDNA library using well known

CC techniques.

XX Sequence 159 BP; 56 A; 23 C; 34 G; 46 T; 0 other;

XX Query Match 83.2%; Score 15.8; DB 24; Length 159;

XX Best Local Similarity 89.5%; Pred. No. 4e+02;

XX Matches 17; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 19:24:08 ; Search time 1455.94 Seconds

(without alignments)
379,791 Million cell updates/sec

Title: US-09-803-736-1079_COPY_13662_13680

Perfect score: 19

Sequence: 1 acaagatacagaacatgc 19

Scoring table: IDENTITY_MOC

Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: gb_ba:*
2: gb_hg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_ma:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_hg_hum:*
31: em_hg_in:*
32: em_hg_juv:*
33: em_hg_mus:*
34: em_hg_other:*
35: em_hg_pln:*
36: em_hg_rod:*
37: em_hg_man:*
38: em_sy:*
39: em_hggo_hum:*
40: em_hggo_mus:*
41: em_hggo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	19	100.0	99814	8	ATF8J2	AL132969 Arabidops
2	18	94.7	153447	9	AC092055	AC092055 Homo sapi
3	18	94.7	185134	9	AC007878	AC007878 Homo sapi
4	18	94.7	200000	9	AP000493	AP000493 Homo sapi
5	17.4	91.6	2479	8	AY087066	AY087066 Arabidops
6	17.4	91.6	7206	1	XYLACRPA	XYLACRPA Scaphylococ
7	17.4	91.6	77432	9	AL358178	AL358178 Homo sapi
8	17.4	91.6	98348	9	AC024933	AC024933 Homo sapi
9	17.4	91.6	105680	8	AC006266	AC006266 Arabidops
10	17.4	91.6	147205	2	AC102991	AC102991 Rattus no
11	17.4	91.6	150946	2	AC117259	AC117259 Mus muscu
12	17.4	91.6	165280	2	AC113475	AC113475 Mus muscu
13	17.4	91.6	166074	2	AC023571	AC023571 Homo sapi
14	17.4	91.6	171059	2	AC121617	AC121617 Rattus no
15	17.4	91.6	173116	2	AC114671	AC114671 Mus muscu
16	17.4	91.6	178116	2	AC073492	AC073492 Homo sapi
17	17.4	91.6	178201	2	AC055807	AC055807 Homo sapi
18	17.4	91.6	178607	2	AC091174	AC091174 Homo sapi
19	17.4	91.6	192963	9	AL354763	AL354763 Homo sapi
20	17.4	91.6	193035	9	AC008180	AC008180 Homo sapi
21	17.4	91.6	194172	2	AC120453	AC120453 Rattus no
22	17.4	91.6	196517	8	ATC0R1Y20	ATC0R1Y20 Arabidops
23	17.4	91.6	230280	2	AC101859	AC101859 Mus muscu
24	17.4	91.6	350000	9	AF130342	AF130342 Homo sapi
25	17	89.5	340	11	HSB286YB1	HSB286YB1 Z53344 H.sapiens
26	17	89.5	30405	9	AL606663	AL606663 Homo sapi
27	17	89.5	50714	2	AL513488	AL513488 Homo sapi
28	17	89.5	81190	2	AC106077	AC106077 Rattus no
29	17	89.5	82315	2	AL592313	AL592313 Homo sapi
30	17	89.5	90571	2	AC106557	AC106557 Rattus no
31	17	89.5	104632	9	AL136124	AL136124 Homo sapi
32	17	89.5	17375	2	AC028207	AC028207 Homo sapi
33	17	89.5	168578	2	AL359756	AL359756 Homo sapi
34	17	89.5	178420	2	AP003548	AP003548 Homo sapi
35	17	89.5	182846	2	AC068223	AC068223 Homo sapi
36	17	89.5	184835	2	AC068091	AC068091 Homo sapi
37	17	89.5	185896	2	AC069564	AC069564 Mus muscu
38	17	89.5	203241	9	AC008282	AC008282 Homo sapi
39	17	89.5	267278	2	AC122252	AC122252 Mus muscu
40	16.4	86.3	336	11	G25146	G25146 human STS E
41	16.4	86.3	390	11	HSJ52F11	HSJ52F11 human STS E
42	16.4	86.3	440	11	G29532	G29532 human STS S
43	16.4	86.3	896	8	AF370475	AF370475 Arabidops
44	16.4	86.3	2192	5	AB072392	AB072392 Agri.stroid
45	16.4	86.3	23137	9	AL670843	AL670843 Human DNA

ALIGNMENTS

RESULT 1
ATF8J2
LOCUS
DEFINITION Arabidopsis thaliana DNA chromosome 3, BAC clone FBJ2.
ACCESSION AL132969
VERSION AL132969.2 GI:7629988
KEYWORDS
SOURCE
ORGANISM
thale cress.
Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE
1 (bases 1 to 97798)
Nykatura,G., Partmann,B., Dauner,D., Sterr,W., Holland,R.,
Weichselgartner,M., Mewes,H.W., Rudd,S., Lemcke,K., Mayer,K.F.X.,

JOURNAL	Unpublished	Quetier,F and Salanoubat,M.
REFERENCE	2 (bases 97548 to 99814)	
AUTHORS	Choisne,N., Robert,C., Brottier,P., Wincker,P., Catolico,L., Attiguenave,F., Saurin,W., Weissensbach,J., Mewes,H.W., Rudd,S., Lemcke,K., Mayer,K.F.X., Quetier,F. and Salanoubat,M.	
JOURNAL	Unpublished	
REFERENCE	3 (bases 1 to 99814)	
AUTHORS	EU Arabidopsis sequencing,project.	
TITLE	Direct Submission	
JOURNAL	Submitted (02-AUG-2000) MIPS, at the Max-Planck-Institut fuer Biochemie, Am Klopferspitz 18a, D-82152 Martinsried, FRG, E-mail: lemcke@mips.biochem.mpg.de,mayer@mips.biochem.mpg.de Project Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France; http://www.genoscope.cns.fr	
COMMENT	On Apr.20, 2000 this sequence version replaced gi:6434218. Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: http://www.mips.biochem.mpg.de/proj/thal/.	
FEATURES	location/Qualifiers	
source	1..99814	
	/organism="Arabidopsis thaliana"	
	/variety="Columbia"	
	/db_xref="taxon:3702"	
	/chromosome="3"	
misc.feature	1..2928	
	/note="this BAC overlaps BAC F3C22"	
gene	4624..9477	
	/gene="F8J2_10"	
CDS	join(4624..4797,5026..5121,5214..5326,5409..5475,5573..5665,5741..5884,6532..6620,6727..6832,6915..6999,7080..7198,7284..7445,7709..7884,7990..8099,8180..8289,8393..8503,8593..8703,9077..9282,9366..9477)	
	/gene="F8J2_10"	
	/note="strong similarity to beta-galactosidase precursor, Carica papaya, EMBL:AF04786	
	Contains Glycosyl hydrolases family 35 putative active site Aa174-186	
	contains EST gb:"T41862, F14142, AA395011, F14066"	
	/codon_start=1	
	/product="beta-galactosidase precursor-like protein"	
	/protein_id="CA86888.1"	
	/db_xref="GI:7529708"	
	/translation="MSMFRKAMILLALICFSLHSTEAAYVYDHKALINGORI LLSGSIHPRSPYEMPDILIRKAKGGGLDTQTYVWNGHEPSPGNYFDODRIDYK TKIVQDALYLDIRIGPYVCAWNGGPGVWLKYVGVAFRTDEPDKLAKMCKFKKI VDMAEAEKLEPTQGGPIILISQLENDYGPWMEGAKAYSKWTAEVALDELSTGVPI MSKQEDAPYPIIDONGPYCEGFKPNSDNPKLMTNMTGWTFEGDALINRPYDIA EHSARIONGSGPKNYMYVGTGNDPRAQVLTATSYDAPIDEYGLIKPEPYSHK ELHAKIKICBPALYSVDPITISLIGKROIHFVSKSCAFLSNYDSARMYRGE PYDLPMSVSTIPDCKREYVYNAKIRAPLITMLIPTSKFMSWESYMGSPSANEAT FYKDSLYQISMTREDKTDYFVFDITIGSDSESLKTGNDPLITIPAGHALHYVNG LKAGSYGALSNSKLIFSQNIKILSVGIKRLALSLVGLPNAGYHTYNTGILCPEPT LKGVNSGPDMSKWKMSYKILRGASMLHFLTAGSSAVKMWIKGVVKKQPLWIKKS FDIPIKNGEPLALDMNTMGGOYVNGVNHNIQRHMPATYANGNGCNAGLYNEKKLS HCCEPSQRMVHYPRSWLKEPQNLLVIEPEGDDPSISLYKAK"	
exon	4624..4797	
	/gene="F8J2_10"	
	/number=1	
intron	4798..5025	
	/gene="F8J2_10"	
	/number=1	
exon	5026..5121	
	/gene="F8J2_10"	
	/number=2	
intron	5122..5213	
	/gene="F8J2_10"	
	/number=2	
exon	5214..5326	
	/gene="F8J2_10"	
	/number=3	
intron	5327..5408	
	/gene="F8J2_10"	
	/number=3	
exon	5409..5475	
	/gene="F8J2_10"	
	/number=4	
intron	5476..5572	
	/gene="F8J2_10"	
	/number=4	
exon	5573..5665	
	/gene="F8J2_10"	
	/number=5	
intron	5666..5740	
	/gene="F8J2_10"	
	/number=5	
exon	5741..5884	
	/gene="F8J2_10"	
	/number=6	
intron	5885..6531	
	/gene="F8J2_10"	
	/number=6	
exon	6532..6620	
	/gene="F8J2_10"	
	/number=7	
intron	6621..6726	
	/gene="F8J2_10"	
	/number=7	
exon	6727..6832	
	/gene="F8J2_10"	
	/number=8	
intron	6833..6914	
	/gene="F8J2_10"	
	/number=8	
exon	6915..6999	
	/gene="F8J2_10"	
	/number=9	
intron	7000..7079	
	/gene="F8J2_10"	
	/number=9	
exon	7080..7198	
	/gene="F8J2_10"	
	/number=10	
intron	7199..7283	
	/gene="F8J2_10"	
	/number=10	
exon	7284..7445	
	/gene="F8J2_10"	
	/number=11	
intron	7446..7708	
	/gene="F8J2_10"	
	/number=11	
exon	7709..7884	
	/gene="F8J2_10"	
	/number=12	
intron	7885..7989	
	/gene="F8J2_10"	
	/number=12	
exon	7990..8099	
	/gene="F8J2_10"	
	/number=13	
intron	8100..8179	
	/gene="F8J2_10"	
	/number=13	
exon	8180..8289	
	/gene="F8J2_10"	
	/number=14	
intron	8290..8392	
	/gene="F8J2_10"	
	/number=14	
exon	8393..8503	
	/gene="F8J2_10"	
	/number=15	
intron	8504..8592	

/gene="F8J2_10"
 /number=15
 8593. .8703
 /gene="F8J2_10"
 /number=16
 8704. .9076
 /gene="F8J2_10"
 /number=16
 9077. .9282
 /gene="F8J2_10"
 /number=17
 9283. .9365
 /gene="F8J2_10"
 /number=17
 9366. .9477
 /gene="F8J2_10"
 /number=18
 15378. .16121
 /gene="F8J2_30"
 complement(15378. .15512,15596. .15711,15902. .16121))
 /gene="F8J2_30"
 complement(15378. .15512,15596. .15711,15902. .16121))
 /gene="F8J2_30"
 /codon_start=1
 /product="putative protein"
 /protein_id="CAB86890.1"
 /db_xref="GI:7529710"
 /translation="MKYQKRPQSSDPSPPDRPGIRSPETSSNONNIEDIMAC
 VTFNALPLCPARELQIDRSPPSHQIDVRHARDPMERAKLDTYMGKEDR
 APSRAESIKDIAVMEELKTKDELIKRWRLFOESQKLVKEQIEKHDELEK"
 complement(15378. .15512)
 /gene="F8J2_30"
 /number=1
 complement(15513. .15595)
 /gene="F8J2_30"
 /number=1
 complement(15596. .15711)
 /gene="F8J2_30"
 /number=2
 complement(15712. .15901)
 /gene="F8J2_30"
 Query Match 100.0%; Score 19; DB 8; Length 99814;
 Best local similarity 100.0%; Pred. No. 49;
 Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ACAAGATACAGACAAATGC 19
 Db 13662 ACAAGATACAGACAAATGC 13680
 RESULT 2
 AC092055 153447 bp DNA linear PRI 30-NOV-2001
 LOCUS Homo sapiens chromosome 3 clone RP11-437K13, complete sequence.
 AC092055 AC010208
 AC092055.2 GI:1155072
 VERSION HTG.
 KEYWORDS Homo sapiens.
 SOURCE Homo sapiens.
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 153447)
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
 Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 153447)
 Kaul,R.K., Olson,M.V., Raymond,C., Clendenning,J., Ivey,R.G. and
 Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (16-JUN-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA

REFERENCE 3 (bases 1 to 153447)
 Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Raymond,C. and
 Haugen,E.D.
 TITLE Direct Submission
 JOURNAL Submitted (30-NOV-2001) Genome Center, University of Washington,
 Box 352145, Seattle, WA 98195, USA
 On Nov 30, 2001 this sequence version replaced gi:14475988.
 COMMENT
 Center: University of Washington Genome Center
 Center Code: UWGC
 Web site: <http://www.genome.washington.edu>
 Contact: uwgchgs@u.washington.edu
 Drafting Center: BCM
 Project Information
 Center project name: chr-3
 Center clone name: RP11-437K13 (bc0423)
 Summary Statistics
 Sequencing vector: M13; 108821; 56% of reads
 Sequencing vector: plasmid; 108752; 44% of reads
 Chemistry: Dye-primer Bodipy; 36% of reads
 Chemistry: Dye-terminator ET; 40% of reads
 Chemistry: Dye-terminator Big Dye; 24% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 153421 bases at least Q40
 Consensus quality: 153447 bases at least Q30
 Insert size: 174533; 14.9% error; agarose-1p
 Insert size: 153451; sum-of-ctrls
 Quality coverage: 8.8x in Q20 bases; agarose-1p
 Quality coverage: 10.0x in Q20 bases; sum-of-ctrls

Overlapping Sequences:
 5': RP11-128H5 (UWGC:bc0242) AC009813
 3': Mapping in progress

Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

This sequence was finished as follows unless otherwise noted:
 all regions were either double-stranded or sequenced with an
 alternate chemistry or covered by high quality data (i.e., Phred
 quality >= 30); an attempt was made to resolve all sequencing
 problems, such as compressions and repeats; all regions were
 covered by at least one plasmid subclone or more than one M13
 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:
 This sequence has been validated by multiple complete digest
 fingerprinting. Comparison of the experimentally derived digest
 fragments with sequence-predicted fragments is given below.
 The electronically-digested sequence consists of both insert and
 vector, in order to accurately represent the entire circular BAC.
 Small fragments below a variable cutoff (approximately 400-800 bp)
 are not resolved in the fingerprint and hence do not appear
 in the table. There are no significant remaining discrepancies
 between the experimental and predicted values. Uniquely ordered
 fragments are separated by dashed lines.

BglII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
4056	4093	8696	9016
2067	2125	6	6382
EcoRI			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
4056	4093	8696	9016
2067	2125	6	6382
HindIII			
SeqDerMap	FngPrnt	SeqDerMap	FngPrnt
4056	4093	8696	9016
2067	2125	6	6382

7045	7091	1685	1659	512	<800
10774	10922	524	<800	449	<800
11209	10922	1697	1659	5446	5446
10805	10922	3748	3745	1502	1469
1999	2125	6738	6848	7223	7271
969	964	3424	3493	5799	5770
4930	4994	1914	1899	5404	5446
627	<800	2342	2345	5006	4967
940	964	6246	6396	3652	3628
375	<800	1226	1214	1221	1204
7831	7827	5688	5807	11172	11215
6819	6800	11968	11673	176	<800
3748	3830	3714	3745	3627	3628
745	768	2124	2124	4677	4558
3020	3142	187	<800	486	<800
3698	3694	10273	10078	5582	5446
1786	1826	2099	2124	3431	3628
4784	4817	2571	2584	4806	4789
249	<800	5826	5807	7877	7874
3225	3320	95	<800	3329	3380
588	<800	3705	3745	3667	3628
1279	1244	637	<800	5779	5770
1808	1826	510	<800	346	<800
6973	7091	4193	4178	9312	9295
119	<800	656	<800	5484	5446
142	<800	1241	1214	3380	3380
2460	2582	686	<800	6112	6120
904	964	508	<800	8210	8238
666	<800	12533	12172	1458	1469
440	<800	21987	22258	572	<800
2996	3142	8507	8423	542	<800
9529	9491	9344	9383	1890	1869
8003	8111	4817	4843	2506	2458
1118	1113	3898	3925	2398	2458
240	<800	6133	6194	694	<800
532	<800			4916	4967

```

FEATURES
  source
    location/Qualifiers
    1. .153447
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="13"
    /clone="RP11-437K13"
    /clone_id="RP11-437K13 human BAC library 11"
BASE COUNT  46396 a 35456 c 33177 g 38418 t
ORIGIN
Query Match          94.7%; Score 18; DB 9; Length 153447;
Best Local Similarity 100.0%; Pred. No. 1.5e+02;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      2 CAAGATACGACGACATGC 19
        |||
Db 151659 CAAGATACGACGACATGC 151676

RESULT 3
AC007878      185134 bp  DNA  linear  PRI 27-Apr-2000
LOCUS
DEFINITION
AC007878      Homo sapiens BAC clone RP11-236P2 from 2, complete sequence.
VERSION
AC007878.2   GI:5732161
KEYWORDS
HTG.
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
JOURNAL
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE
99063792
PUBMED
9847074
REFERENCE
2 (bases 1 to 185134)
Scott, K., Kalicki, J., and Jones, T.
The sequence of Homo sapiens BAC clone RP11-236P2
JOURNAL
Unpublished
3 (bases 1 to 185134)
Waterston, R.H.
Direct Submission
Submitted (19-JUN-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE
4 (bases 1 to 185134)
Waterston, R.H.
Direct Submission
TITLE

```

JOURNAL Submitted (13-AUG-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

REFERENCE 5 (bases 1 to 185134)

AUTHORS Waterston, R.

TITLE Direct Submission

JOURNAL Submitted (27-APR-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA

COMMENT On Aug 13, 1999 this sequence version replaced g1:5103901.

----- Genome Center

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc>

Contact: sapiens@wustl.wustl.edu

----- Summary Statistics

Center project name: H.NH0236P02

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:

Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P.Y., Zhao, B., Frenken, E., Taten, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBAC63.6

NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-236P2; actual end is at base position 185134 of RP11-236P2.

FEATURES

Location/Qualifiers

1..185134

/organism="Homo sapiens"

/db_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone="RP11-236P2"

/clone_lib="RPCI-11"

482..524

/rpt_family="GA-rich"

1363..1586

/rpt_family="MALR"

2455..3592

/rpt_family="L2"

3593..3891

/rpt_family="L2"

3892..3987

/rpt_family="Alu"

4024..4478

/rpt_family="L2"

4502..4872

/rpt_family="MALR"

5280..5575

/rpt_family="L2"

repeat_region

repeat_region

6416..6722

/rpt_family="Alu"

6763..6875

/rpt_family="Alu"

7465..7781

/rpt_family="MER1_type"

7599..7579

/note="Similar to EST AA813754 (NID:g2882439) a16901.s1"

8079..8204

/rpt_family="AT-rich"

8207..8501

/rpt_family="Alu"

9343..9419

/rpt_family="Alu"

9469..9581

/rpt_family="MIR"

9599..9683

/rpt_family="Alu"

9976..10277

/rpt_family="MIR"

10694..10766

/rpt_family="MER2_type"

10794..10908

/rpt_family="Alu"

11244..11280

/rpt_family="MER1_type"

12027..12181

/rpt_family="AT-rich"

12379..12578

/rpt_family="MER1_type"

12580..12597

/note="Similar to EST A1807627 (NID:g5394193) w49b09.x1"

12598..12893

/rpt_family="AT-rich"

12893..13219

/rpt_family="Alu"

13237..13510

/note="Similar to EST A1807627 (NID:g5394193) w49b09.x1"

13759..13797

/rpt_family="Alu"

13827..14045

/rpt_family="AT-rich"

14122..14353

/rpt_family="L1"

14355..14680

/note="Similar to EST AA778964 (NID:g2838295) ab69906.r1"

14681..15189

/rpt_family="Alu"

14685..14860

/note="Similar to EST A1656072 (NID:g4740051) tt43c02.x1"

15132..15366

/note="Similar to EST AA778964 (NID:g2838295) ab69906.r1"

15436..15576

/note="Similar to EST M6900 (NID:g1188066) yy58001.r1"

15494..15874

/rpt_family="L1"

15662..16114

/note="Similar to EST A1688704 (NID:g4899998) wd40e02.x1"

16108..16239

/note="Similar to EST M49591 (NID:g1190757) yy58001.s1"

16749..16884

/rpt_family="Alu"

16926..16984

/rpt_family="L2"

17213..17456

/rpt_family="L2"

17458..17553

/rpt_family="retroviral"

17535..17635

/rpt_family="retroviral"

18147..18253

/rpt_family="retroviral"

repeat_region

```

repeat_region      18446..18488
                    /rpt_family="L2"
repeat_region      18498..18581
                    /rpt_family="L2"
repeat_region      18582..18885
                    /rpt_family="Alu"
repeat_region      18893..19187
                    /rpt_family="Alu"
repeat_region      19188..19242
                    /rpt_family="L2"
repeat_region      19368..19496
                    /rpt_family="MER2_type"
repeat_region      19586..19764
                    /rpt_family="MER2_type"
repeat_region      19798..19892
                    /rpt_family="MER2_type"
repeat_region      20182..20241
                    /rpt_family="L2"
repeat_region      20491..20501
                    /rpt_family="Alu"
repeat_region      20502..20526
                    /rpt_family="ATGn"
repeat_region      20527..20639
                    /rpt_family="Alu"
repeat_region      23226..23514
                    /rpt_family="Alu"

```

Query Match 94.7%: Score 18; DB 9; Length 185134;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR 1 ACAGATACAGACATG 18
 DB 93459 ACAGATACAGACATG 93476

RESULT 4
 LOCUS AP000493/c 200000 bp DNA linear PRI 28-SEP-1999
 DEFINITION Homo sapiens genomic DNA, chromosome 3p21.3, clone:301 to 308,
 anti-oncogene region, section 1/5.
 ACCESSION AP000493
 VERSION AP000493.1 GI:5926660
 KEYWORDS
 ORGANISM Homo sapiens DNA.
 Homo sapiens DNA.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 REFERENCE
 1 (sites)
 Nakamura, Y., Isomura, M., Daigo, Y., Tamari, M. and Ishikawa, S.
 DNA sequence analysis of a 1.2-Mb region on chromosome 3p21.3
 Published only in Database (1999)
 2 (bases 1 to 200000)
 Hirakawa, M., Yamaguchi, H., Imai, K. and Shimada, J.
 Direct Submission
 Submitted (21-SEP-1999) Miki Hirakawa, Japan Science and Technology
 Corporation (JST), Advanced Databases Department, 5-3, Yonbancho,
 Chiyoda-ku, Tokyo 102-0081, Japan (E-mail:mika@tokyo.jst.go.jp,
 URL:http://www-alis.tokyo.jst.go.jp/, Tel:81-3-5214-8491,
 Fax:81-3-5214-8470)
 COMMENT
 This sequence is conducted by Japanese Foundation for Cancer
 Research as a JST sequencing team
 Principal Investigator: Yusuke Nakamura Ph.D
 Phone:+81-3-5449-5372, Fax:+81-3-5449-5433,
 yusuke@cc.ims.u-tokyo.ac.jp
 The sequence is submitted by Human Genome Sequencing in ALIS
 Project of JST
 Japan Science and Technology Corporation (JST)
 5-3, Yonbancho, Chiyoda-ku, Tokyo, 102-0081 Japan
 For further information about this sequences, please visit our
 sequence archive Web site (http://www-alis.tokyo.jst.go.jp/HGS/top.
 html) or send email to webmaster@www-alis.tokyo.jst.go.jp.
 Location/Qualifiers

FEATURES

source

```

1. 200000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3p21.3"
/note="301-308"
complement(41441..41542)
/note="SCC33712:The location is between each flanking site
of PCR primers."
/db_xref="GB:4583716"
complement(67653..67904)
/standard_name="D3S4247"
/note="CHC.ARA22D03.P17192:The location is between each
flanking site of PCR primers."
/db_xref="GB:683838"
76856..77072
/standard_name="D3S1623"
/note="APMD286Yb1:Genethon Marker:The location is between
each flanking site of PCR primers."
/db_xref="GB:610155"
90178..90277
/note="TIGR-A003A11:The location is between each flanking
site of PCR primers."
/db_xref="GB:4586260"
91241..91476
/note="Caalc04:The location is between each flanking site
of PCR primers."
/db_xref="GB:442017"
complement(91341..91539)
/note="TIGR-A003B17:The location is between each flanking
site of PCR primers."
/db_xref="GB:4586314"
complement(91373..91547)
/standard_name="D3S4247"
/note="SHGC-16795:The location is between each flanking
site of PCR primers."
/db_xref="GB:734134"
complement(91438..91565)
/note="WI-18659:The location is between each flanking site
of PCR primers."
/db_xref="GB:4574316"

```

STS

BASE COUNT 52668 a 41602 c 43186 g 62366 t 178 others

ORIGIN

Query Match 94.7%: Score 18; DB 9; Length 200000;
 Best Local Similarity 100.0%; Pred. No. 1.5e+02;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OR 2 CAAGATACAGACATGC 19
 DB 77025 CAAGATACAGACATGC 77008

RESULT 5

LOCUS AY087066/c 2479 bp mRNA linear PIN 26-JUN-2002
 DEFINITION Arabidopsis thaliana clone 3126 mRNA, complete sequence.
 ACCESSION AY087066
 VERSION AY087066.1 GI:21405790
 KEYWORDS
 FLI CDNA.
 SOURCE thale cress.
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
 1 (bases 1 to 2479)
 Haas, B.J., Volkovskiy, N., Town, C.D., Troukhan, M., Alexandrov, N.,
 Feldmann, K.A., Flavell, R.B., White, O. and Salzberg, S.L.
 Full-length messenger RNA sequences greatly improve genome
 annotation
 genome Biol. (2002) In press
 2 (bases 1 to 2479)
 Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and

REFERENCE
 JOURNAL
 2 (bases 1 to 2479)
 Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and

TITLE Feldmann, K.
JOURNAL Full-length cDNA from Arabidopsis thaliana
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 2479)
 Broeyer, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and Feldmann, K.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road, Malibu, CA 90265, USA
COMMENT This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be 5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ms or laer ecotypes and therefore may contain polymorphisms when compared to sequences from Col-0. Genes carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences, selection of clones, and sequence assembly.
FEATURES
 source
 1..2479
 /organism="Arabidopsis thaliana"
 /db_xref="taxon:3702"
 /clone="3126"
 92..2191
 /codon_start=1
 /product="putative glucosyltransferase"
 /protein_id="AA064627.1"
 /db_xref="GI:21592678"
 /translation="MAAPFPMKAKNNNTKGPVYVKKENNNKSWVELESPPHD
 FLVTRTEKSRNKAQULWVLLAKHAKGCTISALPAGTAVKRIAGATDIE
 ISSGVGSLQKONHTKSKSLKVLFWLSLLFELIAVFGWMSGISKLQDQ
 IFNKGFDWYTRWLLRVERLAPLOFLANGCIVLFVSLDRLLICIGFWIRFK
 IKPPKPDISIDSEGNGALFPMVVOIPMCNKEEYQSIQAONLDMKRGILQ
 ILDDSDPTOSLIEEYHMKOKLGAIRYRHRNREGYKAGLSKAMGCVADYEF
 VAIFFADPQPLDPDKTKTIPFQNDNIGVQANFVNKEMLRLINLAFHE
 VEOQNVSPINRFGNGIAGWRIKALPESGMLERTVEMDLIAVRAHLGKMFVL
 NDVCCLEPSTIATKROQRHNSGPMQPLRCLIPAVIKSTISGKKNLIFPL
 RKLLEPYSYFLFCILLPMTEVPEALEPAWVVCILPALSFNLLIPAKSPETVY
 LLENTMSVTKFNNAVSGLFQLSAYEWTYKSGRSEGLDAVDEKTRHQRG
 VSAPEFAKKAETKTKKKKKRRIYMKELSLAFLTLTAATRSLLSAQGIHYELLQ
 GISPLVGLDIDIGOVE"
BASE COUNT 679 a 478 c 543 g 779 t
ORIGIN
 Query Match 91.6%; Score 17.4; DB 8; Length 2479;
 Best Local Similarity 94.7%; Pred. No. 3.4e+02;
 Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ACAAGATACAGACAATGC 19
 Db 687 ACAAGAAACAGACAATGC 665
RESULT 6
SLACRPH 7206 bp DNA linear BCT 06-MAY-1998
DEFINITION Staphylococcus xylosous laetr, laetr, laetr genes and 2 ORF's.
ACCESSION X14599
VERSION X14599.1 GI:2462702
KEYWORDS beta-D-galactosidase; laetr gene; laetr gene; laetr gene; lactose operon; lactose permease; transcription activator.
SOURCE Staphylococcus xylosous.
ORGANISM Bacteria; Firmicutes; Bacillales; Staphylococcus.
REFERENCE 1 (bases 1 to 7206)
AUTHORS Bassias, J. and Bruckner, R.

TITLE Regulation of lactose utilization genes in Staphylococcus xylosous
JOURNAL J. Bacteriol. 180 (9), 2273-2279 (1998)
MEDLINE 98233718
PUBMED 9573174
REFERENCE 2 (bases 1 to 7206)
AUTHORS Bruckner, R.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-1997) R. Bruckner, Mikrobielle Genetik, Universitaet Tuebingen, Auf der Koenigsstelle 28, D-72076 Tuebingen, FRG
FEATURES
 source
 1..7206
 /organism="Staphylococcus xylosous"
 /strain="DSM20267"
 /isolate="cured strain Caa"
 /db_xref="taxon:1288"
 <1..814
 /note="ORF1"
 /codon_start=2
 /transl_table=1
 /product="transcriptional regulator from the lysR-type"
 /protein_id="CAA74934.1"
 /db_xref="GI:2462703"
 /db_xref="SWISS-PROT:033812"
 /translation="LAEVKAQSPFKAEVNIYSPSISRDRIKLENDYDVYKFEKHK
 SKMTLTSQGEKLYQYVQRYVLEQALRODLMONNVAAGDLKGSFPEGRYRISQOL
 TKLAQMYPELHIHVLNDSRTIVQIKNNIVDQVIEKKIQQNNAIISTPLAQDEIVLI
 KRKSSLSNLEFCIFREGSGTRVQENGLNLSINPLVIVNNTSLIKSVHNGNFS
 IYKSTLISDLEQLEVINDIREFRYLLIKKXIDEKKRIVISVAKQWNE"
 complement(954..1864)
 /gene="laetr"
 /complement(954..1828)
 /gene="laetr"
 /product="lactose operon transcription activator"
 /note="major"
 /evidence="experimental"
 complement(954..1815)
 /gene="laetr"
 /product="lactose operon transcription activator"
 /note="minor"
 /evidence="experimental"
 complement(954..1793)
 /gene="laetr"
 /codon_start=1
 /evidence="experimental"
 /transl_table=1
 /product="lactose operon transcription activator"
 /protein_id="CAA74935.1"
 /db_xref="SWISS-PROT:033813"
 /translation="MOYIMKKRQKLIIDANLAEGGIEIGVPMGYNTVFKSKVLHV
 TGSEGFSTYAGETVHLFADGIIPLERGMGEVYKRSNNPTVYVYMGNGQILNIYSR
 CSTVDSHYLQDPTDITKNIQITCKLSQIESNNSDILNIQYLYLYTLIDKRPK
 IFVQVQDVNEDIQHADVFTNIYQKHITFDVAKSVNITRSHLTKIKRKNCGSKRE
 YLYIMYHASQLILHRTSLIDISIKOVYKDPILSKNFTKHFISASEYRHHFSIN
 NKQ"
 complement(1800..1805)
 /gene="laetr"
 /complement(1824..1829)
 /gene="laetr"
 /note="promoter 1"
 /complement(1835..1840)
 /gene="laetr"
 /note="promoter 2"
 complement(1847..1852)
 /gene="laetr"
 /note="promoter 1"
 complement(1859..1864)
 /gene="laetr"
 /note="promoter 2"
 join(1885..1914,1921..6419)
 /gene="laetr"
 complement(1885..1890)

```

/gene="laccp"
1909..1914
/gene="laccp"
1921..6419
misc_feature
1927..1940
evidence=experimental
/gene="laccp"
1955..1961
note="cfe element"
/gene="laccp"
1955..1961
/gene="laccp"
1967..3355
/gene="laccp"
1967..3355
/gene="laccp"
1967..3355
/codon_start=1
evidence=experimental
/transl_table=1
/product="lactose permease"
/protein_id="CA47436.1"
/db_xref="GI:2462705"
/transl_table="NNRBLTGKQKAFGAFGALGKAIENIVSPLMEFYITDIYVGHSPA
FYGVWLVFVARIIMDAINDPIMGIVDNTNNNGKFEKWTYIGTLINAVYTLFTNMDL
SOTAMYIYISITISYMGWYIMMDIPYWSLPMTHDRREBELSVIPREFASIAFT
VTRGLPFIHKLDIPGSGSDSTGIPPAICSLIPIETICVYKPEKPEOEMKELIG
IKVFKDGRILPRKNEELIAGVLEFNLCIDTNGSITTYEVVNAHELPAIFNS
MILCEWGLILPRFKWGRIAENISIIISGLIILISGLIAGSIAOSWTIILIGAG
ILRISGPMWITVSLADVIDYGVKFGGNSIITSNTFLKTSQAVAAIIVGLG
LSILGTPENASQSVITINGLIRIMLVSPVIFCIAMFIHKAFLMKGDEFITDIETLQ
YKROFERERIK"
3356..3361
/gene="laccp"
3356..3361
/gene="laccp"
3370..6354
/gene="laccp"
3370..6354
/gene="laccp"
3370..6354
/codon_start=1
evidence=experimental
/transl_table=1
/product="beta-D-galactosidase"
/protein_id="CA47437.1"
/db_xref="GI:2462706"
/transl_table="MKOOFHNELEVOHNLPRAFYIPYQOGBADYHETRYNTN
ATLNGEYRYPEPEEATLNSKOSKTLVPSWNLGYDOYLNTQYPIPNP
PVPKDNPCGHTTRKTTIDETQOYDHLNPEYDASAFWNNPEFISQISHAISE
TINSDLNAQIDIKLIRAHNLKSVETLNPGEVNASISVNHODVANHMLST
ENPVLTYILITDEVITQVIGIRVAINNOFYINGOSIKLGTNHDHSHPTGVG
TESHEKDLLELMKQGNALFTAHYKSPLEETMOYFYFVMSADIEHPTGVG
EDNNEFNIADDSKEPILIRIASIMPLKNSIYSVSLGSGKMGKAGAR
AKSIDNPEPIYEGTLYRDKOHHYLSIMISMSPSESEETYSNDLKPILL
CEYAHAMNSPGDIHAOTLVAEQVDSPTGVMWMDHAIDYGMKDNQPIFRYGGFG
EKILHGNFCVDGIVFENRVPHEGYERKOEHRPLINYSOEDRYLRQUDPIPAKY
MEVEATVNLNGKGTISEIPLSNLPHTAQITDIDSYINQHSIVLIRKLTIDIT
RIENFELGHDOIVORTLKEONEBQSDTEILLVYDKLIVYSGKSTYVNDAS
LESVLKHHIYISONTNNINRAPDNDNTINKDMAYSGYKIDITRVHDOIVENDE
VSLIENIMNDVAPVYLGTVYVHVOGNGLNLYYDLERDKAPYLRGLGLTLPK
APROKYYKGPSPSYODKGVANVYLDGCTVNDGEIHRIRQEGTSGHNETTVESID
GCKKVIYISDNTFPTNTHSLKOLFTTHKDALFEPDQIYIIDYAGSGISGNSGCP
ELNEAFRLNHRHTESENLKFLV
6380..6419
terminator
/gene="laccp"
note="rho independent"
complement(6476..7117)
/gene="orf5"
complement(6476..7105)
/gene="orf5"

```

```

/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="CA474938.1"
/db_xref="GI:2462707"
/db_xref="SPTREML:033816"
/transl_table="MKGLIISYSSKSTYTYLIVAILAVAFINPACIPLML
FLTPVDNIKHKDSKMYITISLPSGRSAVNSYSPFGITLITIGTIFVSVSILIT
TQDFNVILMSIIVGIGAGYAMPEPFRKPEPNSVNLITSTFVLFFVFRPM
NLSRVAVSGSDAINDFNSIYVGVVAVGLVIAFLVSYVSLKIFRNOEL"
complement(7111..7117)
/gene="orf5"
BASE COUNT 2560 a 1064 c 1248 g 2334 t
ORIGIN
Query Match 91.6%; Score 17.4; DB 1; Length 7206;
Best Local Similarity 94.7%; Pred. No. 3.3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ACAAGATACGACACATGC 19
DB 420 AAAAGATACGACACATGC 438
RESULT 7
AL358178/c 77432 bp DNA linear PRI 15-DEC-2000
LOCUS Human DNA sequence from clone Rp11-812120 on chromosome 6, complete
DEFINITION
ACCESSION AL358178
VERSION AL358178.14 GI:11876053
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 77432)
REFERENCE
AUTHORS Skuce, C.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-2000) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerquest@sanger.ac.uk
COMMENT
On Dec 17, 2000 this sequence version replaced gi:11602546.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence has been finished according to sequence map criteria
as follows. An attempt is made to resolve all sequencing problems,
such as compressions and repeats, but not necessarily within known
annotated repeat sequence elements. Where the sequence is
ambiguous, there is an annotation using the 'unsure' feature key.
The following abbreviations are used to associate primary accession
numbers given in the feature table with their source databases:
Em, EMBL; Sw, SWISSPROT; Tr, TrEMBL; Wp, WormEP; Information
on the WormEP database can be found at
http://www.sanger.ac.uk/Projects/C_elegans/wormep
This sequence was generated from part of bacterial clone configs of human
chromosome 6, constructed by the Sanger Centre Chromosome 6 Mapping
Group. Further information can be found at
http://www.sanger.ac.uk/HGP/Chr6
Rp11-812120 is from the library RPI1-11.3 constructed by the group
of Pieter de Jong. For further details see
http://www.chori.org/bacpac/home.htm
VECTOR: pBAC3.6
IMPORTANT: This sequence is not the entire insert of clone
Rp11-812120. It may be shorter because we sequence overlapping
sections only once, except for a 100 base overlap.
The true left end of clone Rp11-812120 is at 1 in this sequence.
The true left end of clone Rp11-55004 is at 77333 in this sequence.
The true right end of clone Rp11-385F7 is at 50771 in this

```


Query Match	91.6%	Score 17.4	DB 9	Length 77432
Best Local Similarity	94.7%	Pred. No. 3	Idet: 02	
Matches 18	Conservative 0	Mismatches 1	Indels 0	Gaps 0
OY 1	ACAAAGATACAGAACATTC 19			
Db 25342	ACAAAGATACAGAACATTC 25324			
RESULT 8				
LOCUS AC024933				
DEFINITION Homo sapiens 3 BAC RP11-219D15 (Roswell Park Cancer Institute Human				
ACCESSION AC024933				
VERSION AC024933.33	GI:21535842			
SOURCE HTG.				
ORGANISM human.				
human sapiens				
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
1 (bases 1 to 98348)				
Mzany,D.M., Adams,C., Adio-Ogunola,B., Ali-osman,F.R., Allen,C.,				
Alshrooks,S.L., Amarunge,H.C., Are,J.R., Ayale,M., Banks,T.,				
Barber,J., Benton,J., Birmage,K., Blankenburg,K., Bonnin,D.,				
Bouch,J., Boyle,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,				
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,				
Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,				
Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,				
Cleveland,C.D., Cox,C., Coyle,M., Dathorne,S.R., David,R.,				
Devila,M.L., Davis,C., Davy-Carroll,L., Dedechter,D.A.,				
DeJamey,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,				
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,				
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,				
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Franz,P.,				
Gabibi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,				
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,				
Harris,C., Harris,K., Hart,M., Haylak,P., Hayes,A., He,X.,				
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holmway,C.,				
Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J.,				
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,				
Joudes,S., Karlsson,E., Kelly,S., Khan,U., King,L., Koreah,J.,				
Kovar,C., Krtovic,J., Kucharski,A., Landry,N., Leal,B., Lewis,L.C.,				
Lewis,L., Liu,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,				
Loulsgeed,H., Lozano,R.J., Lu,X., Luster,A., Luster,R., Luna,R.,				
Ma,J., Maheshwari,M., Mapa,P., Martin,R., Martindale,A.,				
Mathine,E., Massey,E., Mawhney,E., McLeod,M.P., Medori,M.,				
Mel,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabadi,K.,				
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,				
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,				
Nickerson,E., Nwoketwo,S., Ogun,N., Okunnu,G., Ocranu,N.,				
Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,				
Pickens,R., Primus,E., Pu,L.L., Ralls,M., Ren,Y., Rives,M.,				
Rojas,A., Rojibkhan,I., Rolfe,M., Ruiz,S., Sarey,G., Scherer,S.,				
Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,E.,				
Scralke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,				
Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,				
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,				
Vera,Y., Villalona,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,				
Warren,R., Washington,C., Wallington,S., Williams,G.,				
Williamson,A., Wlezyk,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,				
Wu,Y.F., Zhou,J., Zorrilla,S., Zorrilla,S.L., Weinstock,G., and				
Gibbs,R.				
Direct Submission				
Unpublished				

REFERENCE	2 (pases 1 to 98348)
AUTHORS	Morley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (03-MAR-2000) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	3 (pases 1 to 98348)
AUTHORS	Morley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (17-JUN-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
REFERENCE	4 (pases 1 to 98348)
AUTHORS	Morley,K.C.
TITLE	Direct Submission
JOURNAL	Submitted (21-JUN-2002) Human Genome Sequencing Center, Department

CLONE LENGTH This sequence does not necessarily represent the entire insert of this clone. Overlapping regions of clones are only sequenced and submitted once, so the sequence for the remainder of the insert may be found in the record for the adjacent clones. Overlapping clones are noted at the beginning and end of the Features listing.

ANNOTATION OF FEATURES:
STSs are identified using ePCR (Genome Res. 7:541-550) searches of a local database that includes entries from dbSTS, GDB, and local mapping efforts.
Repeats are identified using RepeatMasker (A. Smit and P. Green, unpublished.) for Human and Mouse sequences.
Genes and Region of sequence similarity are identified by BLAST (Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the EST and cDNA sequences. Genes demonstrate at least two exons flanked by consensus splice sites that maintained sequence continuity across the splice junctions. Sequences that are not identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum standard of double strand coverage with a minimum of 2 clones and 2 reads with no ambiguities or 2 chemistries with a minimum of 2 clones and 3 reads with no ambiguities. If the sequence quality for a region does not meet this standard, it will be indicated in the annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality standards - estimated error rate less than 1 per 10,000 bases. Reports of lowest quality individual bases and measures of base quality are listed below. Description of the metrics can be found at URL: <http://www.hgsc.bcm.tmc.edu:8088/quality.info/gcnbank.annotation.html>.

```

FEATURES
source
    Location/Qualifiers
    1..98348
    /organism="Homo sapiens"
    /db_xref="taxon:9606"
    /chromosome="3"
    /clone="RP11-219D15"
    complement(608..881)
    /rpt_family="L2"
    complement(892..1142)
    /rpt_family="LIMB3"
    complement(1157..1463)
    /rpt_family="AluXs"
    complement(1540..2615)
    /rpt_family="LIMB6"
    complement(2623..3350)
    /rpt_family="L2"
    3381..3446
    /rpt_family="MIR"

```

```

repeat_region complement(3947..4112)
/rpt_family="MIR"
repeat_region 5074..5158
/rpt_family="(GGCA)n"
repeat_region 5195..5230
/rpt_family="(GGA)n"
repeat_region 5454..5605
/rpt_family="MIR"
repeat_region 5672..5871
/rpt_family="MIR"
repeat_region complement(5897..5984)
/rpt_family="L2"
repeat_region complement(6224..6287)
/rpt_family="MIR"
repeat_region 7280..7389
/rpt_family="MLTK"
repeat_region 7390..7698
/rpt_family="AluSc"
repeat_region 7699..7915
/rpt_family="MLTK"
repeat_region 8117..8146
/rpt_family="(CAAA)n"
repeat_region complement(8183..8303)
/rpt_family="L2"
repeat_region complement(8810..8873)
/rpt_family="L2"
repeat_region complement(9532..9663)
/rpt_family="MIR"
repeat_region complement(10399..10461)
/rpt_family="MIR"
repeat_region 10503..10793
/rpt_family="AluSc"
repeat_region complement(10813..10967)
/rpt_family="MIR"
repeat_region complement(10991..11126)
/rpt_family="MIR"
repeat_region 11577..11611
/rpt_family="MIR"
repeat_region 12354..12547
/rpt_family="MER63A"
repeat_region 12550..12773
/rpt_family="MER69"
repeat_region 12770..12901
/rpt_family="MER69"
repeat_region 13165..13243
/rpt_family="L2"
repeat_region complement(13247..14052)
/rpt_family="LIM4"
repeat_region 15123..15490
/rpt_family="HEIC"
repeat_region complement(15789..15851)
/rpt_family="MIR"
repeat_region 16173..16216
/rpt_family="(CA)n"
repeat_region 16618..16775
/rpt_family="MIR"
repeat_region 17014..17198
/rpt_family="MLTJ"
repeat_region 17200..17598
/rpt_family="LIME2"
repeat_region 17601..17628
/rpt_family="AT_rich"
repeat_region 18326..18397
/rpt_family="(CA)n"
repeat_region 18436..18816
/rpt_family="L2"
STS 18935..19036
/standard_name="73095"
repeat_region 18961..18992
/rpt_family="(CA)n"
repeat_region 19068..19490
/rpt_family="MLTIF"
repeat_region 20798..20820

```

```

Query Match          91.6%; Score 17.4; DB 9; Length 98348;
Best Local Similarity 94.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ACACATACAGAAATATGC 19
DB 9341 ACACATACAGAAATATGC 9359
|||||

```

```

RESULT 9
AC006266 105680 bp DNA linear pL1 25-FEB-1999
LOCUS Arabidopsis thaliana BAC FIK3 from Chromosome IV near 21 cm,
DEFINITION complete sequence.
AC006266
VERSION AC006266.1 GI:4092472
KEYWORDS
SOURCE
ORGANISM

```

```

REFERENCE
AUTHORS
Nascimento L., Huang E.N., de la Bastide M., Habermann K.,
See L.H., Preston R., Matero A., Rodriguez M.A., O'Shaughnessy A.,
Vil M.D., Shekter M., Shady I., Shah R., Spiegel L.A., Schultz K.,
Parnell L.D., Dedhia N.N. and McCombie W.R.
Arabidopsis thaliana BAC FIK3 from Chromosome IV near 21 cm
Unpublished
2 (bases 1 to 105680)
Nascimento L., Huang E.N., de la Bastide M., Habermann K.,
See L.H., Preston R., Matero A., Rodriguez M.A., O'Shaughnessy A.,
Vil M.D., Shekter M., Shady I., Shah R., Spiegel L.A., Schultz K.,
Parnell L.D., Dedhia N.N. and McCombie W.R.
Direct Submission
Submitted (01-JAN-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724, USA
3 (bases 1 to 105680)
Parnell L.D. and McCombie W.R.
Direct Submission
Submitted (20-FEB-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724
Arabidopsis thaliana BAC FIK3 from Chromosome IV, near 21 cm
4 (bases 1 to 105680)
Parnell L.D. and McCombie W.R.
Direct Submission
Submitted (25-FEB-1999) Lita Annenberg Hazen Genome Sequencing
Center, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring
Harbor, NY 11724

```

```

REMARK
AUTHORS
JOURNAL
COMMENT
REMARK
AUTHORS
JOURNAL
COMMENT

```

Arabidopsis thaliana BAC FIK3 from Chromosome IV, near 21 cm
 BAC FIK3 maps to near 21 cm on the Lister & Dean RI map. Position 1
 of FIK3, the 97 end, is oriented toward the NOR and the telomere
 and position 105680 is oriented toward the long arm. For more
 information on the mapping, sequencing and annotation of FIK3,
 please see <http://www.cshl.org/arabidopsis/FLK3-tillepage.html>. A
 graphic view of our annotation is also available at this url. Gene
 models are built with exons predicted by GENSCAN
 (<http://CCR-081.mit.edu/GENSCAN.html>), MZEF
 (<http://www.cshl.org/genefinder>) and GRAIL
 (<http://combio.ornl.gov/tools/index.shtml>) and with splice sites
 predicted by NetPlantGene
 (<http://www.cbs.dtu.dk/netgene/cbsnetgene.html>). Alternate exons
 not used in building the gene models are presented on the web pages
 associated with FIK3. Genes are numbered according to the scheme
 BAC.gene.number. Typically, these numbers progress from 1 upwards
 as one moves from position 1 of the BAC. Protein sequences encoded
 by the genes are assigned to a functional category with the aid of
 similarity searches and comparison to the Prosite
 (<http://expasy.hcuge.ch/prosite/prosite.html>) and Pfam
 (<http://pfam.wustl.edu/>) libraries. A description of these

categories can be found at
http://muntjac.mips.biochem.mpg.de/Arabid/. Genomic repeats are
typically located by TBLASTX analysis or with RepeatMasker and an
attempt is made to classify the function of each repeat as either
transposon, putative microsatellite, LINE, direct repeat,
centromeric repeat, etc.

If you have any questions or confirmatory or contradictory evidence
concerning the annotation of FIK3, please direct email to Larry
Parnell at parnell@cshl.org.

At position 1 we have left a minimal overlap of 200 bp with T24G23.

```
FEATURES
Source          Location/Qualifiers
1..105680
   /organism="Arabidopsis thaliana"
   /cultivar="Columbia"
   /db_xref="taxon:3702"
   /chromosome="IV"
   /map="near 21 CM"
repeat_region   1..2510
   /note="function=transposon; transposon_type=Athlia"
   /rpl_type=dispersed
misc_feature    1..200
   /note="overlap with T24G23 from position 105221 to 105420"
repeat_region   2666..2849
   /note="function=180_bp_repeat; H33 type of Athlia 180 bp
   repeat"
   /rpl_type=dispersed
repeat_region   2850..6557
   /note="function=putative transposon; similar to MJE4,
   Genbank accession number AB013393; similar to T26C24,
   Genbank accession number AC004705; similar to T9E19,
   Genbank accession number AF104920"
   /rpl_type=dispersed
repeat_region   3669..3791
   /note="function=direct_repeat; cognate A of direct repeat
   from position 3669 to 4038"
   /rpl_type=direct
repeat_region   3795..3917
   /note="function=direct_repeat; cognate B of direct repeat
   from position 3669 to 4038"
   /rpl_type=direct
repeat_region   3921..4038
   /note="function=direct_repeat; cognate C of direct repeat
   from position 3669 to 4038"
   /rpl_type=direct
repeat_region   10568..10782
   /note="function=putative LINE; similar to T5123, Genbank
   accession number AC002983"
   /rpl_type=dispersed
gene            13380..14816
   /gene="FIK3.1"
   /note="encodes hypothetical protein; gene model last
   edited on 23 Feb 99"
   /evidence=not_experimental
   join(13380..13412,13508..13551,14159..14258,14413..14420,
   14582..14816)
   /gene="FIK3.1"
   /note="functional catalog ID=99"
   /codon_start=1
   /evidence=not_experimental
   /product="hypothetical protein"
   /protein_id="AAD15480.1"
   /db_xref="GI:4309696"
   /translation="MPEDSNMRCKEWPGRKRSQRIELNMLGGDTLYAMTTEERK
   KGPQILDKTKATLAKELGAGVSRKRRPSDPLDOKLEAKFAKKRNALVA
   ESAGSRNADNEDDESESKSAFGKKRKNSTPH"
   18751..21382
   /gene="FIK3.2"
   /note="encodes putative DNA-directed RNA polymerase;
   genomic copy of EST T46243; gene model last edited on 23
   Feb 99"
```

```
5'UTR
/evidence=not_experimental
join(18751..18858,19594..19616)
/gene="FIK3.2"
/gene="5'-UTR based on a comparison to EST T46243"
join(19617..19812,20093..20217)
/gene="FIK3.2"
/note="similar to A. fulgidus DNA-directed RNA polymerase
subunit M, Genbank accession number AE001019; similar to
S. pombe DNA-directed RNA polymerase III, Genbank
accession number O13896; functional catalog ID=04.05.01"
/codon_start=1
/evidence=not_experimental
/product="putative DNA-directed RNA polymerase subunit"
/protein_id="AAD15481.1"
/db_xref="GI:4309697"
/translation="MERCPCGNLLRYRGGSSRPFCGYPVANIERRVEIKKOLL
VRSIEPVVTKDDIPYAAETEARPCRGCHKAYFKSMQIRSADEPESFYRLCKEPT
WRE"
complement(23335..25821)
/gene="FIK3.3"
/note="encodes putative glucosyltransferase; gene model
last edited on 23 Feb 99"
/evidence=not_experimental
complement(join(23335..23958,24110..24223,24316..24606,
24673..24981,25060..25821))
/gene="FIK3.3"
/note="functional catalog ID=01.05"
/codon_start=1
/evidence=not_experimental
/product="putative glucosyltransferase"
/protein_id="AAD15482.1"
/db_xref="GI:4309698"
/translation="MAPKEPMWAKGNNNNRFGKTPVYVKNENMNMWVLESPSHD
FLVTRHESRNKNAROLTWLILAKARAGCCTGLSALFAGTAVRARIARLQIE
ISSGSGLOKONTKSKLFYSCLKFLAMSTILGFEIAYFKGSGFSGLOQF
IKKGFEPGMYTARWVLLRYEYLAAPLQFLANGCIYELVOSLDRILIGCGPIRTRK
IKVYKPPSISLDSGNGGAFLEPVIQNIQDPCNREKYOQSLAANCINMPCXKILIO
IIDSDDPTOSLIEKVEYHKQKLRARIYHRRNRGKAKSNANCSYKQDFE
VAIFDADPQLPDLTKTIPHRKDNBEIGLQARWBSVKNKEELTRIONINLAFPE
VQGVAVSYPLNFGPNTAGVWIKALDSGSGMLERYVBDMLIVRHLHGKVFLE
NDVECCCELPESEYARQOHRWHSQWQLRLCLPAVLKKSISGKFFNLPLFPL
RKLIPEYSEFLFCIILPMTFVPEAEALPAMVVCYIPATMSPLNLPAPKFFETVY
LLEPNTMSVTENFMAVSGLPOLGSAYEVWVKKSGRSSEGLAALVEDEKTYHVG
VSAPEAKKAEKTKRKKRKNRYKELSLAPLITATSLSLAAGIHFFELLFO
GISFLLIGDLIGQVE"
29655..35800
/gene="FIK3.4"
/note="similar to A. thaliana hypothetical proteins"
/pseudo
/evidence=not_experimental
38121..40108
/gene="FIK3.5"
/note="encodes hypothetical protein; gene model last edited
on 24 Feb 99"
/evidence=not_experimental
join(38121..38131,39066..39192,39485..39733,39815..39905,
39981..40108)
/gene="FIK3.5"
/note="similar to A. thaliana hypothetical proteins;
functional catalog ID=99"
/codon_start=1
/evidence=not_experimental
/product="hypothetical protein"
/protein_id="AAD15483.1"
/db_xref="GI:4309699"
/translation="WVRGHSINNDTEKQELIPKQGEKHEEGKEEKEEKK
Query Match 91.6%; Score 17.4; DB 8; Length 105680;
Best Local Similarity 94.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
1 ACAAGTACGAACAATGC 19
||||| |||||||||
```

Db 25226 ACAAGAAAGACATGC 25244

RESULT 10

AC102991

LOCUS

AC102991

DEFINITION

Rattus norvegicus clone CH230-58N23, *** SEQUENCING IN PROGRESS

ACCESSION

AC102991

VERSION

AC102991.3

KEYWORDS

HTG: HTGS: PHASEL

SOURCE

Norway rat.

ORGANISM

Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

REFERENCE

1 (bases 1 to 147205)

AUTHORS

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amarantune,H.C., Are,J.R., Ayale,M., Banks,T., Barthala,J., Benton,J., Bimaye,K., Blankenburg,K., Bonini,D., Bouck,J., Bowe,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P., Buiay,C., Burch,P., Burkett,C., Butrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,B., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flaga,N., Ford,J., Foster,P., Franz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Correll,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,P., Hayes,A., Hernandez,J., Hernandez,O., Hodson,A., Hognes,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,B., Kelly,S., Khan,U., King,L., Korah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichlargo,O., Lien,C., Liu,J., Liu,W., Lollisged,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Matthei,E., Massey,E., Mayhew,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M., Minter,G., Miner,Z., Mitchell,T., Mohabhat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Neelam,S., Neelam,S., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogutu,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojuckan,I., Rolfe,M., Ruiz,S., Savary,G., Schermer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I., Sodergren,F., Sonalike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A., Taber,P., Tameria,A., Tameria,K., Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S., Tsumaki,K., Vasquez,L., Vera,Y., Villalobos,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S., Williams,G., Williamson,A., Wlezyk,R., Woodson,S., Worley,K., Wu,C., Wu,Y., Wu,Y.F., Zorrilla,S., Nelson,D., Weinstock,G., and Gibbs,R.

TITLE

Unpublished

REFERENCE

2 (bases 1 to 147205)

AUTHORS

Worley,K.C.

JOURNAL

Submitted (24-NOV-2001)

TITLE

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 11, 2002 this sequence version replaced g1:17973122.

Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
Project Information
Center project name: CH230-58N23
Center clone name: CH230-58N23
Sequencing vector: Plasmid
Chemistry: Dye-terminator Big Dye 1.00% of reads
Assembly program: Phrap, version 0.990329
Consensus quality: 89364 bases at least Q40
Consensus quality: 93104 bases at least Q20
Consensus quality: 95864 bases at least Q20

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a "working draft" sequence. It currently consists of 67 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1	1445:	contig of 1445 bp in length
1446	1545:	gap of unknown length
1546	2641:	contig of 1096 bp in length
2642	2742:	gap of unknown length
2742	3801:	contig of 1060 bp in length
3802	3902:	gap of unknown length
3902	5329:	contig of 1428 bp in length
5330	5429:	gap of unknown length
5430	6469:	contig of 1040 bp in length
6470	6569:	gap of unknown length
6570	7639:	contig of 1070 bp in length
7640	7739:	gap of unknown length
7740	8764:	contig of 1025 bp in length
8765	8864:	gap of unknown length
8865	10179:	contig of 1315 bp in length
10180	10279:	gap of unknown length
10280	11530:	contig of 1251 bp in length
11531	11630:	gap of unknown length
11631	12762:	contig of 1133 bp in length
12763	12862:	gap of unknown length
12863	14503:	contig of 1641 bp in length
14504	14603:	gap of unknown length
14604	16202:	contig of 1599 bp in length
16203	16302:	gap of unknown length
16303	18230:	contig of 1928 bp in length
18231	18330:	gap of unknown length
18331	19474:	contig of 1144 bp in length
19475	19574:	gap of unknown length
19575	20856:	contig of 1282 bp in length
20857	20956:	gap of unknown length
20957	22723:	contig of 1767 bp in length
22724	22823:	gap of unknown length
22824	24201:	contig of 1378 bp in length
24202	24301:	gap of unknown length
24302	25698:	contig of 1397 bp in length
25699	25798:	gap of unknown length
25799	27477:	contig of 1679 bp in length
27478	27577:	gap of unknown length
27578	28753:	contig of 1176 bp in length
28754	28853:	gap of unknown length
28854	30630:	contig of 1777 bp in length
30631	30730:	gap of unknown length
30731	31836:	contig of 1106 bp in length
31837	31936:	gap of unknown length
31937	33432:	contig of 1496 bp in length
33433	33532:	gap of unknown length
33533	35148:	contig of 1616 bp in length
35149	35248:	gap of unknown length

```

* 35249 37457: contig of 2209 bp in length
* 37458 37557: gap of unknown length
* 37558 39030: contig of 1473 bp in length
* 39031 39130: gap of unknown length
* 39131 40729: contig of 1599 bp in length
* 40730 40829: gap of unknown length
* 40830 42544: contig of 1715 bp in length
* 42545 42644: gap of unknown length
* 42645 44325: contig of 1681 bp in length
* 44326 44425: gap of unknown length
* 44426 45870: contig of 1445 bp in length
* 45871 45970: gap of unknown length
* 45971 47667: contig of 1797 bp in length
* 47668 47867: gap of unknown length
* 47868 49502: contig of 1635 bp in length
* 49503 49602: gap of unknown length
* 49603 52253: contig of 2651 bp in length
* 52254 52353: gap of unknown length
* 52354 53952: contig of 1599 bp in length
* 53953 54052: gap of unknown length
* 54053 55950: contig of 1898 bp in length
* 55951 56050: gap of unknown length
* 56051 58082: contig of 2032 bp in length
* 58083 58182: gap of unknown length
* 58183 59801: contig of 1619 bp in length
* 59802 59901: gap of unknown length
* 59902 62459: contig of 2568 bp in length
* 62470 62569: gap of unknown length
* 62570 64332: contig of 1763 bp in length
* 64333 64432: gap of unknown length
* 64433 66430: contig of 1998 bp in length
* 66431 66530: gap of unknown length
* 66531 68353: contig of 1823 bp in length
* 68354 68453: gap of unknown length
* 68454 70587: contig of 2134 bp in length
* 70588 70687: gap of unknown length
* 70688 71952: contig of 1265 bp in length
* 71953 72052: gap of unknown length
* 72053 74472: contig of 2420 bp in length
* 74473 74572: gap of unknown length
* 74573 76172: contig of 1600 bp in length
* 76173 76272: gap of unknown length
* 76273 78331: contig of 2059 bp in length
* 78332 78432: gap of unknown length
* 78433 79716: contig of 1285 bp in length
* 79717 79816: gap of unknown length
* 79817 81234: contig of 1418 bp in length
* 81235 81334: gap of unknown length
* 81335 84468: contig of 3134 bp in length
* 84469 84568: gap of unknown length
* 84569 85700: contig of 1132 bp in length
* 85701 85800: gap of unknown length
* 85801 88631: contig of 2831 bp in length
* 88632 88731: gap of unknown length
* 88732 90755: contig of 2024 bp in length
* 90756 90855: gap of unknown length
* 90856 93179: contig of 2324 bp in length
* 93180 93279: gap of unknown length
* 93280 96584: contig of 3305 bp in length

```

```

Query Match      91.6%; Score 17.4; DB 2; Length 147205;
Best Local Similarity 94.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

OY 1 ACAAGATACGACAAATGC 19
    ||||| ||||| |||||
Db 57271 ACAAGATCAGACAAATGC 57289

```

```

RESULT 11
LOCUS AC117259 150956 bp DNA linear HTG 25-JUN-2002
DEFINITION Mus musculus chromosome UNK clone RP24-492019, WORKING DRAFT

```

```

ACCESSION      SEQUENCE, 6 unordered pieces.
AC117259
VERSION        AC117259.2 GI:21553274
KEYWORDS       HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE         house mouse.
ORGANISM       Mus musculus.
REFERENCE      Eukaryotic Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eultheria; Rodentia; Sciurognathi; Muridae; Mus.
               1 (bases 1 to 150956)
               McPherson,J.D. and Waterston,R.H.
               The sequence of Mus musculus clone
               Unpublished
               2 (bases 1 to 150956)
               McPherson,J.D. and Waterston,R.H.
               Direct Submission
               Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park
               Parkway, St. Louis, MO 63108, USA
               3 (bases 1 to 150956)
               McPherson,J.D. and Waterston,R.H.
               Direct Submission
               Submitted (25-JUN-2002) Genome Sequencing Center, 4444 Forest Park
               Parkway, St. LOUIS, MO 63108, USA
               On Jun 25, 2002 this sequence version replaced gi:20069847.
COMMENT

```

```

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@wustl.wustl.edu
----- Project Information -----
Center project name: M_BB0492019

```

```

----- Summary Statistics -----
Sequencing vector: M13; 0%
Sequencing vector: plasmid; 100%
Chemistry: Dye-terminator Big Dye; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 148911 bases at least Q40
Consensus quality: 149538 bases at least Q30
Consensus quality: 149980 bases at least Q20
Insert size: 161000; agarose-fp
Insert size: 150456; sum-of-contigs
Quality coverage: 9.58 in Q20 bases; agarose-fp
Quality coverage: 10.27 in Q20 bases; sum-of-contigs

```

```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 6 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

```

1 1245: contig of 1245 bp in length
* 1246 1345: gap of unknown length
* 1346 2415: contig of 1070 bp in length
* 2416 2516: gap of unknown length
* 2516 3688: contig of 1173 bp in length
* 3688 3789: gap of unknown length
* 3789 4898: contig of 1110 bp in length
* 4898 4998: gap of unknown length
* 4998 46659: contig of 41661 bp in length
* 46660 46759: gap of unknown length
* 46760 150956: contig of 104197 bp in length.

```

```

FEATURES
source
1..150956
/organism="Mus musculus"
/db_xref="taxon:10090"
/chromosome="UNK"
/clone="RP24-492019"

```

```

misc_feature
1..1245
/note="assembly_name:Contig26"

```

```

misc_feature      1346..2415
                    /note="assembly_name:Contig40"
misc_feature      2516..3688
                    /note="assembly_name:Contig43"
misc_feature      3789..4898
                    /note="assembly_name:Contig45"
misc_feature      4999..46659
                    /note="assembly_name:Contig48"
misc_feature      46760..150956
                    /note="assembly_name:Contig49"
BASE COUNT      46895 a 28915 c 28762 g 45884 t      500 others
ORIGIN
Query Match      91.6%; Score 17.4; DB 2; Length 150956;
Best Local Similarity 94.7%; Pred. No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACAGATACAGACAGACATGC 19
        ||||| ||||| ||||| |||||
Db      5087 ACAGACAGACAGACATGC 5105

RESULT 12
AC113475      165250 bp      DNA      linear      HTG 06-JUN-2002
LOCUS      Mus musculus clone RP23-304B1, WORKING DRAFT SEQUENCE, 14 ordered
DEFINITION      pieces.
AC113475      AC113475.3 GI:21327503
VERSION      HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
KEYWORDS      house mouse.
SOURCE      Mus musculus
ORGANISM      Mus musculus
REFERENCE      1 (bases 1 to 165250)
AUTHORS      Birren, B., Linton, L., Nusbaum, C. and Lander, E.
JOURNAL
REFERENCE
AUTHORS

TITLE
JOURNAL
REFERENCE
AUTHORS

Submitted (01-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165250)
Direct Submission
Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhalter, B.,
Boukhalter, B., Brown, A., Camarata, J., Campopiano, A., Chang, J.,
Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collamore, A.,
Cook, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J., Dodge, S.,
Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gaidyna, S.,
Gande, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
Kamat, A., Karatas, A., Kells, C., Laroque, K., Lamazares, R.,
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,
McDwan, P., McKernan, K., Meldrum, J., Meneus, L., Milova, T.,
Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,
Peterson, K., Phunhkhang, P., Pierre, N., Pollard, V., Raymond, C.,
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,
Rosetti, M., Roy, A., Santos, R., Schauback, R., Seaman, S.,
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., Young, G.,
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

```

```

TITLE
JOURNAL
COMMENT

Johnson, R., Jones, C., Kamat, A., Karatas, A., Kells, C., Laroque, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N.,
Matthews, C., McCarthy, M., McDwan, P., McKernan, K., Meldrum, J.,
Meneus, L., Milova, T., Mlenga, Y., Murphy, T., Naylor, J., Nguyen, C.,
Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P.,
O'Neill, D., Oliver, J., Peterson, K., Phunhkhang, P., Pierre, N.,
Pollard, V., Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C.,
Rogov, P., Roman, J., Rosetti, M., Roy, A., Santos, R., Schauback, R.,
Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N.,
Strauss, N., Subramanian, A., Talamas, J., Testaye, S.,
Theodore, J., Topham, K., Travers, M., Travis, N., Trigilio, J.,
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.,
Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (06-JUN-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 6, 2002 this sequence version replaced gi:21313891.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RV/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L23705
Center clone name: 304.B-1
----- Summary Statistics
Sequencing vector: plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 159392 bases at least Q40
Consensus quality: 161973 bases at least Q40
Consensus quality: 162935 bases at least Q20
Insert size: 17200; agarose-fp
Insert size: 163950; sum-of-ctrls
Quality coverage: 5.8 in Q20 bases; agarose-fp
Quality coverage: 6.0 in Q20 bases; sum-of-ctrls
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 14 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
1      1340: contig of 1340 bp in length
*      1341 1440: gap of 100 bp
*      1441 2023: contig of 583 bp in length
*      2024 2123: gap of 100 bp
*      2124 3204: contig of 1081 bp in length
*      3205 3304: gap of 100 bp
*      3305 4737: contig of 1433 bp in length
*      4738 4837: gap of 100 bp
*      4838 7717: contig of 2880 bp in length
*      7718 7817: gap of 100 bp
*      7818 9993: contig of 2176 bp in length
*      9994 10093: gap of 100 bp
*      10094 13995: contig of 3902 bp in length
*      13996 14095: gap of 100 bp
*      14096 18493: contig of 4398 bp in length
*      18494 18593: gap of 100 bp
*      18594 22106: contig of 3513 bp in length
*      22107 22206: gap of 100 bp
*      22207 38119: contig of 15913 bp in length
*      38120 38219: gap of 100 bp
*      38220 57734: contig of 19515 bp in length
*      57735 57834: gap of 100 bp
*      57835 88306: contig of 30472 bp in length

```

FEATURES

* 88307 88406: gap of 100 bp

* 88407 127689: contig of 39283 bp in length

* 127690 127789: gap of 100 bp

* 127790 165250: contig of 37461 bp in length.

Location/Qualifiers

1. 165250

/organism="Mus musculus"

/db_xref="taxon:10090"

/clone="Rp23-304B1"

/clone_id="RP23-23 Female Mouse BAC"

1. 1340

/note="assembly_fragment"

clone_end:SP6

vector_side:left

misc_feature 1441..2023

/note="assembly_fragment"

2124..3204

/note="assembly_fragment"

3305..4737

/note="assembly_fragment"

4838..7717

/note="assembly_fragment"

7818..9993

/note="assembly_fragment"

10094..13995

/note="assembly_fragment"

14096..18493

/note="assembly_fragment"

18594..22106

/note="assembly_fragment"

22207..38119

/note="assembly_fragment"

38220..57734

/note="assembly_fragment"

57835..88306

/note="assembly_fragment"

88407..127689

/note="assembly_fragment"

127790..165250

/note="assembly_fragment"

misc_feature 51665 a 30524 c 30865 g 50896 t 1300 others

BASE COUNT

ORIGIN

Query Match 91.6%; Score 17.4; DB 2; Length 165250;

Best Local Similarity 94.7%; Pred. No. 3.1e+02;

Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ACAAGTACAGACATGC 19

||||| |||||||||

Db 164070 ACAAGACAGACATGC 164088

RESUME 13

AC023571/c 166074 bp DNA linear HTG 24-AUG-2002

LOCUS Homo sapiens chromosome 6 clone RP11-623N6 map 6, WORKING DRAFT

DEFINITION

SEQUENCE 21 unordered pieces.

AC023571

AC023571.2 GI:7210021

VERSION

KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

1 (bases 1 to 166074)

Birten, B., Linton, L., Nusbaum, C. and Lander, E.

Homo sapiens chromosome 6, clone RP11-623N6

Unpublished

2 (bases 1 to 166074)

Birten, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,

Anderson, S., Baldwin, O., Barna, N., Bede, F., Boguslavsky, L.,

Boukhalter, B., Brown, A., Burkett, G., Campoliano, A., Castle, A.,

Choepe, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P.,

REFERENCE

AUTHORS

JOURNAL

TITLE

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

Deatellano, K., Dewar, K., Dodge, S., Domino, M., Doyle, M.,

Renestor, C., Ferreira, P., Fitzhugh, W., Forrest, C., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaton, A., Horton, L.,

Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, B., Karatas, A.,

Klein, J., Landers, T., Laroque, K., Lehoczy, J., Levine, R.,

Iliev, C., Liu, G., Locke, K., MacDonald, P., Marquis, N., McCarthy, M.,

McKernan, P., McGurk, A., McKernan, C., Mienga, Y., Morrow, J.,

Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, T. M.,

Peterson, K., Pierre, N., Pisan, C., Pollara, V., Raymond, C.,

Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S.,

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,

Subramanian, A., Talamas, J., Testa, S., Theodore, J., Tirrell, A.,

Travers, M., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B.,

Wu, X., Wyman, D., Ye, W. J., Young, C., Zainoun, J., Zimmer, A. and

Zody, M.

Direct Submission

Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 166074)

Birten, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N.,

Anderson, S., Baldwin, O., Barna, N., Bede, F.,

Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G.,

Campoliano, A., Castle, A., Choepe, Y., Colangelo, M., Collins, S.,

Collymore, A., Cooke, P., Deatellano, K., Dewar, K., Diaz, J. S.,

Dodge, S., Domino, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,

Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,

Grand-Pierre, N., Grant, G., Hagos, B., Heaton, A., Horton, L.,

Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, B., Karatas, A.,

Klein, J., Laroque, K., Lamazares, R., Landers, T., Lehoczy, J.,

Levine, R., Liu, G., Locke, K., MacDonald, P., Marquis, N.,

McCarthy, M., McKernan, P., McGurk, A., McKernan, C., Mienga, Y., Morrow, J.,

Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Neill, P.,

O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,

Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,

Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,

Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,

Testa, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,

Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,

Young, C., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

On Mar 8, 2000 this sequence version replaced gi:6978283.

All repeats were identified using RepeatMasker:

Smit, A. F. A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: MIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

----- Project Information

Center Project name: 16781

Center Clone name: 623_N_6

----- Summary Statistics

Sequencing vector: M13; 77815, 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 147055 bases at least Q40

Consensus quality: 156779 bases at least Q30

Consensus quality: 161144 bases at least Q20

Insert size: 170000; agarose-fp

Insert size: 164074; sum-of-ctrls

Quality coverage: 3.5 in Q20 bases; agarose-fp

Quality coverage: 3.6 in Q20 bases; sum-of-ctrls

NOTE: This is a 'working draft' sequence. It currently

consists of 21 contigs. The true order of the pieces

is not known and their order in this sequence record is

arbitrary. Gaps between the contigs are represented as


```

* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1057: contig of 1057 bp in length
* 1058 1157: gap of 100 bp
* 1158 2801: contig of 1644 bp in length
* 2802 2901: gap of 100 bp
* 2902 5488: contig of 2587 bp in length
* 5489 5588: gap of 100 bp
* 5589 7107: contig of 1519 bp in length
* 7108 7207: gap of 100 bp
* 7208 8646: contig of 1439 bp in length
* 8647 8746: gap of 100 bp
* 8747 12635: contig of 3889 bp in length
* 12636 12735: gap of 100 bp
* 12736 16425: contig of 3690 bp in length
* 16426 16525: gap of 100 bp
* 16526 20266: contig of 3741 bp in length
* 20267 20366: gap of 100 bp
* 20367 25270: contig of 4904 bp in length
* 25271 25370: gap of 100 bp
* 25371 31863: contig of 6499 bp in length
* 31870 31969: gap of 100 bp
* 31970 36772: contig of 4803 bp in length
* 36773 36872: gap of 100 bp
* 36873 43127: contig of 6255 bp in length
* 43128 43227: gap of 100 bp
* 43228 50522: contig of 7295 bp in length
* 50523 50622: gap of 100 bp
* 50623 58189: contig of 7567 bp in length
* 58190 58289: gap of 100 bp
* 58290 68122: contig of 9833 bp in length
* 68123 68222: gap of 100 bp
* 68223 80517: contig of 12295 bp in length
* 80518 80617: gap of 100 bp
* 80618 92897: contig of 12280 bp in length
* 92898 92997: gap of 100 bp
* 92998 108700: contig of 15703 bp in length
* 108701 108800: gap of 100 bp
* 108801 126942: contig of 18142 bp in length
* 126943 127042: gap of 100 bp
* 127043 145386: contig of 18344 bp in length
* 145387 145486: gap of 100 bp
* 145487 160704: contig of 20588 bp in length.
*
FEATURES
Source
1. 166074
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="6"
/map="6"
/clone="RP11-623N6"
/clone_lib="RP11 Human Male BAC"
1. 1057
/note="assembly_fragment"
misc_feature
1158. 2801
/note="assembly_fragment"
misc_feature
2902. 5488
/note="assembly_fragment"
misc_feature
5589. 7107
/note="assembly_fragment"
clone_end:77
vector_side:left
misc_feature
7208. 8646
/note="assembly_fragment"
clone_end:SP6
vector_side:left
misc_feature
8747. 12635
/note="assembly_fragment"
misc_feature
12736. 16425
/note="assembly_fragment"
misc_feature
16526. 20266
/note="assembly_fragment"

```

```

misc_feature 20367. 25270
/note="assembly_fragment"
misc_feature 25371. 31869
/note="assembly_fragment"
misc_feature 31970. 36772
/note="assembly_fragment"
misc_feature 36873. 43127
/note="assembly_fragment"
misc_feature 43228. 50522
/note="assembly_fragment"
misc_feature 50623. 58189
/note="assembly_fragment"
misc_feature 58290. 68122
/note="assembly_fragment"
misc_feature 68223. 80517
/note="assembly_fragment"
misc_feature 80618. 92897
/note="assembly_fragment"
misc_feature 92998. 108700
/note="assembly_fragment"
misc_feature 108801. 126942
/note="assembly_fragment"
misc_feature 127043. 145386
/note="assembly_fragment"
misc_feature 145487. 166074
/note="assembly_fragment"
BASE COUNT 52217 a 30395 c 30737 g 50721 t 2004 others
ORIGIN

```

```

Query Match 91.6%; Score 17.4; DB 2; Length 166074;
Best Local Similarity 94.7%; Pred.No. 3.1e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ACAGATACGACGACATGC 19
Db 131482 ACAGATACGACGACATGC 131464
|||||
AC121617/c
AC121617/c
LOCUS
DEFINITION
Rattus norvegicus clone CH230-240M15, *** SEQUENCING IN PROGRESS
ACCESSION
AC121617
VERSION
AC121617.3 GI:21746668
KEYWORDS
HTG; HTGS; PHASE1.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 171059)

```

```

REFERENCE
1 (bases 1 to 171059)
AUTHORS
Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alstbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayale,M., Banks,T.,
Barbieri,U., Benton,J., Blmage,K., Blankenburg,K., Bonnin,D.,
Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,
Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
Carroll,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dim,B.H.,
Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J.,
Hernandez,O., Hodgson,A., Hogues,M., Holikoway,C., Hollins,B.,
Homs,J.F., Howard,S., Huber,J., Huliy,S., Hume,U., Jackson,L.E.,
Jacobsen,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
Karlsom,E., Kelly,S., Khan,U., King,L., Koryab,J., Kovac,C.,
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,

```



```

FEATURES
  source
    1. 171059
    /organism="Rattus norvegicus"
    /db_xref="taxon:10116"
    /clone="CH230-248M15"

BASE COUNT      44893 a 36765 c 35645 g 47457 t      6299 others

ORIGIN
Query Match      91.6%; Score 17.4; DB 2; Length 171059;
Best Local Similarity 94.7%; Pred. No. 3e+02;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 ACAAGATACGAGACATGTC 19
        ||||| ||||| |||||
Db 44832 ACAAGATCAGACATGTC 44814

RESULT 15
AC114671/c
LOCUS      173116 bp      DNA      linear      HTG 10-JUN-2002
DEFINITION Mus musculus clone RP24-199D21, WORKING DRAFT SEQUENCE, 8 ordered
            pieces.
AC114671
VERSION     AC114671.3 GI:21362151
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      house mouse.
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 173116)
AUTHORS     Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
            Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
            Bouckhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
            Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
            Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
            Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
            Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
            Kanat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
            Landers, T., Lehoczek, J., Levine, R., Lindblad-Toh, K., Liu, G.,
            Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
            McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
            Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
            Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
            Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V.,
            Raymond, C., Retta, R., Riedack, M., Riley, R., Rise, C., Rogoy, P.,
            Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
            Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Scojanovic, N.,
            Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
            Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
            Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.
DIRECT SUBMISSION
Submitted (10-MAR-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
JOURNAL
  3 (bases 1 to 173116)
REFERENCE
  Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,
  Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L.,
  Bouckhalter, B., Brown, A., Camarata, J., Campolano, A., Chang, J.,
  Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
  Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
  Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
  Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,
  Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
  Kanat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R.,
  Landers, T., Lehoczek, J., Levine, R., Lindblad-Toh, K., Liu, G.,
  Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
  McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L.,
  Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
  Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D.,
  Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V.,
  Raymond, C., Retta, R., Riedack, M., Riley, R., Rise, C., Rogoy, P.,
  Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
  Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Scojanovic, N.,
  Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
  Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,
  Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,
  Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

```

TITLE JOURNAL COMMENT

Galagan, J., Gardyna, S., Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanat, A., Karatas, A., Kells, C., Lacroque, K., Lamazares, R., Landers, T., Lehoczek, J., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., Meldrum, J., Menus, L., Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Nordu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riedack, M., Riley, R., Rise, C., Rogoy, P., Roman, J., Roselli, M., Roy, A., Santos, R., Schauer, S., Schnupack, R., Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J., Zemdek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (10-JUN-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 10, 2002 this sequence version replaced g1.21328554.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIRB
Web site: <http://www-seq.wi.mit.edu>
Contact: sequence_submissions@genome.wi.mit.edu
Project information

Center project name: 199_D-21
Center clone name: 124616
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye, 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 170646 bases at least Q40
Consensus quality: 171597 bases at least Q30
Consensus quality: 172061 bases at least Q20
Insert size: 16800; agarose-ff
Insert coverage: 9.1 in Q20 bases; agarose-ff
Quality coverage: 8.8 in Q20 bases; sum-of-coverage

NOTE: This is a 'working draft' sequence. It currently consists of 8 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the pieces of the gaps between them are based on estimates that have been provided by the submitter.
This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.
1 2673: contig of 2673 bp in length
2674 2773: gap of 100 bp
2774 2773: contig of 4739 bp in length
2774 7612: gap of 100 bp
7613 7612: contig of 7393 bp in length
7613 15005: contig of 19564 bp in length
15006 15105: gap of 100 bp
15106 34669: contig of 19564 bp in length
34670 34769: gap of 100 bp
34770 55824: contig of 21055 bp in length
55825 55924: gap of 100 bp
55925 55924: contig of 29284 bp in length
55925 85308: gap of 100 bp
85309 85308: contig of 40674 bp in length
85309 125982: gap of 100 bp
125983 126082: contig of 47034 bp in length.
126083 173116: contig of 47034 bp in length.

FEATURES
source
1. 173116
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RP24-199D21"
/clone_lib="RPCT-24 Male Mouse BAC"

```

misc_feature      1..2673
                  /note="assembly-fragment"
misc_feature      2774..7512
                  /note="assembly-fragment"
misc_feature      7613..15005
                  /note="assembly-fragment"
misc_feature      15106..34669
                  /note="assembly-fragment"
misc_feature      34770..55824
                  /note="assembly-fragment"
misc_feature      55925..85208
                  /note="assembly-fragment"
misc_feature      85309..125982
                  /note="assembly-fragment"
misc_feature      126083..173116
                  /note="assembly-fragment"
BASE COUNT      46302 a 39948 c 40972 g 45194 t      700 others
ORIGIN

```

```

Query Match      91.6%; Score 17.4; DB 2; Length 173116;
Best Local Similarity 94.7%; Pred. No. 3e+02; 1; Indels 0; Gaps 0;
Matches 18; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY      1 ACAAGATACAGAACAAATGC 19
        | ||||| ||||| |||||
Db 122118 ATAGATACAGAACAAATGC 122100

```

Search completed: January 23, 2003, 21:25:00
 Job time : 1617.94 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 20:33:53 ; Search time 1161.6 Seconds

(without alignments)
223.078 Million cell updates/sec

Title: US-09-803-736-1022_COPY_10615_10630

Perfect score: 16
Sequence: 1 ttgtgtgcaattttc 16

Scoring table: IDENTITY_MTC

Gapop 10.0, Gapext 1.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Listing first 45 summaries

```

1: em_estba:*
2: em_esthnm:*
3: em_estlin:*
4: em_estlmu:*
5: em_estlov:*
6: em_estlpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estlun:*
16: em_estlom:*
17: gb_gss:*
18: em_gss_hnm:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vtl:*
22: em_gss_fun:*
23: em_gss_man:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rtd:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	100.0	390	9	AA875684	TENDU0156
2	100.0	415	13	B1142716	B1142716 RK73a03.Y
3	100.0	466	13	BM290045	BM290045 EST576579
4	100.0	548	17	A2152204	A2152204 SP.0009.A
5	100.0	551	17	B97310	B97310 R33A15TR.TA
6	100.0	662	17	BH842853	BH842853 TC3-57A4.

Result No.	Score	Query Match	Length	ID	Description
7	100.0	926	12	BE960519	BE960519 601653267
8	93.8	132	13	BG948370	BG948370 IP1.12.H0
9	93.8	200	10	AM480273	AM480273 30583 MAR
10	93.8	218	9	AF049574	AF049574 AF049574
11	93.8	251	14	B0556095	B0556095 HA038A08-
12	93.8	267	9	AL731053	AL731053 AL731053
13	93.8	267	10	BB113077	BB113077 BB113077
14	93.8	267	13	B1054222	B1054222 B1054222
15	93.8	285	9	AV142630	AV142630 AV142630
16	93.8	306	13	B1141338	B1141338 IP1.54.D1
17	93.8	324	12	B1269430	B1269430 NF015F01
18	93.8	324	10	BG727927	BG727927 f079f09.Y
19	93.8	327	10	BB188316	BB188316 BB188316
20	93.8	337	14	B0595949	B0595949 PFES03B2
21	93.8	344	17	AQ722178	AQ722178 HS-5201.B
22	93.8	389	10	BE029444	BE029444 K28C07.Y
23	93.8	399	17	B38870	B38870 HS-1048-B1-
24	93.8	411	14	BQ090010	BQ090010 rc58612.Y
25	93.8	441	17	AQ115116	AQ115116 RC111-54
26	93.8	448	10	BE578694	BE578694 RK01B01.Y
27	93.8	461	9	AA780825	AA780825 ac66h07.s
28	93.8	464	10	BE224216	BE224216 K28C06.Y
29	93.8	464	12	BG688339	BG688339 335919.BA
30	93.8	478	10	AM933390	AM933390 EST59329
31	93.8	490	12	BE517076	BE517076 NXS1.008-
32	93.8	492	17	AQ483145	AQ483145 RC11-1-2
33	93.8	501	10	AV948748	AV948748 AV948748
34	93.8	502	12	BE716408	BE716408 sa21905.Y
35	93.8	503	12	BF325321	BF325321 su32b09.Y
36	93.8	505	14	BQ079501	BQ079501 sm14e10.Y
37	93.8	513	17	AZ902960	AZ902960 RC11-24-1
38	93.8	523	10	BB697318	BB697318 BB697318
39	93.8	524	17	AQ31085	AQ31085 RC11-1-3
40	93.8	534	17	TA31611P	TA31611P TA31611P
41	93.8	538	17	BH403784	BH403784 AC-ND-101
42	93.8	543	17	A2412362	A2412362 IK0185H17
43	93.8	544	14	BQ780752	BQ780752 UT-R-PRO-
44	93.8	547	10	BB692544	BB692544 BB692544
45	93.8	551	10	BE331504	BE331504 sp14d07.Y

ALIGNMENTS

RESULT 1
AA875684/c
LOCUS
DEFINITION
AA875684 390 bp mRNA linear EST 21-SEP-2000
TENDU0156 F. cruzi epimastigote normalized cDNA library Trypanosoma
cruzi cDNA clone Sp11 3', mRNA sequence.

ACCESSION
AA875684.1 GI:2981514

VERSION
AA875684.1

KEYWORDS
EST.

SOURCE
Trypanosoma cruzi.

ORGANISM
Trypanosoma cruzi.

REFERENCE
1 (bases 1 to 390)
Trypanosoma, Schizotrypanum.

AUTHORS
Porcel, B.M., Tran, A.-N., Tammi, M., Nyarady, Z., Rydaker, M., Urményi, T.P., Rondinelli, E., Petersson, B., Andersson, B. and Aslund, L.

TITLE
Gene survey of the pathogenic protozoan Trypanosoma cruzi

JOURNAL
Genome Res. 10 (8), 1103-1107 (2000)

MEDLINE
20414748

COMMENT
Contact: Aslund L
Department of Medical Genetics
Uppsala University
Biomedical Center, Box 589, S-751 23 Uppsala, Sweden
Tel: 46 18 471 45 85
Fax: 46 18 52 68 49
Email: lena.aslund@medgen.uu.se
Seq primer: M13 Forward.
Location/Qualifiers
1..390
/organism="Trypanosoma cruzi"

Query Match 100.0%; Score 16; DB 9; Length 390;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 42 TTGTGTCATTTTC 16
 |||
 |||

RESULT 2
 LOCUS B1142716 415 bp mRNA linear EST 10-JUL-2001
 DEFINITION tk73a03.y1 Meloidogyne javanica egg PAMPI v6 Chiapelli McCarter
 B1142716
 Meloidogyne javanica cDNA 5', mRNA sequence.
 B1142716.1 GI:14624426
 EST.
 root-knot nematode.
 Meloidogyne javanica
 Eukaryota; Metazoa; Nematoda; Chromadorea; Tylenchida; Tylenchina;
 Tylenchoidea; Heteroderidae; Meloidogyneae; Meloidogyne.
 1 (bases 1 to 415)
 McCarter, J., Clifton, S., Chiapelli, B., Pape, D., Martin, J., Wylie, T.,
 Dante, M., Merritt, M., Hillier, L., Kucaba, T., Theising, B., Bowers, Y.,
 Gibbons, M., Ritter, E., Bennett, J., Franklin, C., Tsagaris, H., R.,
 Konko, I., Kennedy, S., Maguire, L., Beck, C., Underwood, K., Steptoe,
 M., Allen, M., Person, B., Swaller, T., Harvey, N., Schurk, R., Kohn, S.,
 Shin, T., Jackson, Y., Cardenas, K., McCann, R., Waterston, R. and
 Wilson, R.
 The Washington Univ. Nematode EST Project, 1999
 Unpublished (1999)
 Contact: McCarter JP
 The Washington Univ. Nematode EST Project, 1999
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 The library was constructed by Brandt Chiapelli and Dr. James
 McCarter (bchiapeli@watson.wustl.edu & jmcarter@watson.wustl.edu) at
 Washington University, St. Louis. DNA Sequencing by: Washington
 University Genome Sequencing Center St. Louis.
 Seq primer: -40RP from Glibco
 High quality sequence stop: 364.
 Location/Qualifiers
 1. 415
 /organism="Meloidogyne javanica"
 /db_xref="taxon:6303"
 /clone_lib="Meloidogyne javanica egg PAMPI v6 Chiapelli
 McCarter"
 /dev_stage="enriched for eggs"
 /lab_host="DH10B"
 /note="Vector: PAMPI (Glibco): The library was constructed
 by Brandt Chiapelli and Dr. James McCarter at Washington
 University, St. Louis. The cDNA was made by using Dynabead
 oligo-dT priming (Dyna). PCR based library using a
 modified protocol from the SMART PCR cDNA Synthesis kit
 from Clontech. Directionally cloned into the DUG sites of
 PAMPI. Nematodes were provided by Dr. David Bird of North
 Carolina State University."

BASE COUNT 153 a 49 c 45 g 167 t 1 others
 ORIGIN

Query Match 100.0%; Score 16; DB 13; Length 415;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 160 TTGTGTCATTTTC 16
 |||
 |||

RESULT 3
 LOCUS BM290045 466 bp mRNA linear EST 01-JUL-2002
 DEFINITION ESM576579 AVSG Amblyomma variegatum cDNA clone AVA121 5' end, mRNA
 sequence.
 BM290045
 ACCESSION BM290045.1 GI:21640012
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Amblyomma variegatum.
 Amblyomma variegatum
 Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;
 Parasitiformes; Ixodida; Ixodidae; Amblyomma.
 1 (bases 1 to 466)
 Nene, V., Lee, Y., Skilton, R., Quackenbush, J., Gardner, M. and Bishop
 A partial gene index of ESTs from the salivary glands of Amblyomma
 variegatum
 Unpublished (2002)
 Contact: Vish Nene
 Parasite Genomics Group
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: 301-610-5968
 Fax: 301-838-0208
 Email: nene@igf.org
 Seq primer: M13 reverse.
 Location/Qualifiers
 1. 466
 /organism="Amblyomma variegatum"
 /db_xref="taxon:34610"
 /clone_lib="AVSG"
 /clone_id="AVSG"
 /tissue_type="Salivary glands"
 /dev_stage="Adult"
 /note="Vector: PCWVSPORT6.0.ccd; Salivary glands were
 dissected on day five after initiation of feeding. Total
 RNA was prepared using acid guanidium
 thiocyanate-phenol-chloroform extraction. The cDNA library
 was custom prepared by Invitrogen Corporation. Briefly,
 first strand cDNA was primed using oligo(dT) containing a
 NotI site. Size fractionated double stranded cDNA was
 ligated to EcoRV-NotI cleaved vector and electroporated
 into E.coli."

BASE COUNT 119 a 93 c 113 g 141 t
 ORIGIN

Query Match 100.0%; Score 16; DB 13; Length 466;
 Best Local Similarity 100.0%; Pred. No. 4.8e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 105 TTGTGTCATTTTC 120
 |||
 |||

RESULT 4
 LOCUS AZ152204 548 bp DNA linear GSS 28-AUG-2000
 DEFINITION SP_0009_A2_H10_T7 Strongyloides purpuratus, purple sea urchin,
 sperm genomic BAC library Strongylocentrotus purpuratus genomic
 clone Plate-9 Col=20 Row=O, DNA sequence.
 AZ152204

ACCESSION

VERSION	A2152204.1 GI:8304105		
KEYWORDS	GSS.		
SOURCE	Strongylocentrotus purpuratus		
ORGANISM	Strongylocentrotus purpuratus		
	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;		
	Echinoidea; Euechinoidea; Echinacea; Echinoida;		
	Strongylocentrotidae; Strongylocidae.		
REFERENCE	1 (bases 1 to 548)		
AUTHORS	Cameron,R.A., Malais,G., Rast,I.T.P., Martinez,P., Biondi,T.R., Swartzell,S., Wallace,J.C., Pouska,A.J., Livingston,B.T., Wray,G.A., Etlenson,C.A., Lehrach,H., Bittlen,R.J., Davidson,E.H. and Hood,L.		
TITLE	A sea urchin genome project: Sequence scan, virtual map, and additional resources		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (17), 9514-9518 (2000)		
MEDLINE	2040256		
COMMENT	Contact: Cameron, RA, Davidson, EH, Hood, L Division Of Biology 156-29 California Institute of Technology Pasadena California 91125, USA Tel: (626) 395-8421 Fax: (626) 793-3047 Email: acameron@caltech.edu Plate: 9 row: 0 column: 20 Seq primer: 77 Class: BAC ends High quality sequence stop: 548. Location/Qualifiers 1..548 /organism="Strongylocentrotus purpuratus" /db_xref="taxon:7668" /clone="plate:9 Col=20 Row=0" /clone_1lb="Strongylocentrotus purpuratus, purple sea urchin, sperm genomic BAC library" /note="Organ: sperm; Vector: BACe3.6; BAC clones in E-coli DH10B"		
BASE COUNT	170 a	120 c	89 g 166 t 3 others
ORIGIN			
Query Match	100.0%;	Score 16;	DB 17; Length 548;
Best Local Similarity	100.0%;	Pred. No. 4,8e+03;	
Matches	16;	Conservative 0;	Mismatches 0; Indels 0; Gaps 0;
QY	1 TTGTGTCATTTTC 16 		
Db	536 TTGTCATTTTC 521		
RESULT 5			
B97310/c			
LOCUS			
DEFINITION	T33A15RF TAND Arabidopsis thaliana genomic clone T33A15, DNA sequence.		
ACCESSION	B97310		
VERSION	B97310.1		
KEYWORDS	GI:299289		
SOURCE	GSS.		
ORGANISM	thale cress.		
	Arabidopsis thaliana		
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
	Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.		
REFERENCE	1 (bases 1 to 511)		
AUTHORS	Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Adams,M.D. and Venter,J.C.		
TITLE	A BAC End Sequence Database for Identifying Minimal Overlaps in Arabidopsis Genomic Sequencing, update 3		
JOURNAL	Unpublished (1997)		
COMMENT	Other-GSSs: T33A15RF Contact: Steve Rounsley Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA		

```

Tel: 301 838 0200
Fax: 301 838 0208
Email: ronsley@tigr.org
Seq primer: M15 Reverse
Class: BAC ends
High quality sequence stop: 551.

FEATURES
    source
        location/Qualifiers
            1..551
                /organism="Arabidopsis thaliana"
                /strain="Columbia"
                /db_xref="taxon:3702"
                /clone="t33a15"
                /clone_id="FRAMU"
                /sex="hermaphrodite"
                /note="Vector: BelosAC11, Site_1: HindIII, Site_2: HindIII
                ; Produced by Rod Wang"
                149 a          148 g          156 t

BASE COUNT
    149 a          148 g          156 t

ORIGIN
    Query Match
    Best Local Similarity 100.0%; Score 16; DB 17; Length 551;
    Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

    1 TTTGTTGTCATTTTTC 16
    |||||||
    Db 302 TTCTTGTCATTTTTC 287

RESULT 6
    BH842853/c 662 bp DNA linear GSS 13-JUN-2002
    LOCUS BH842853/c
    DEFINITION TC3-57A4, TV TC3 Trypanosoma cruzi genomic clone TC3-57A4, DNA
    sequence.
    ACCESSION BH842853
    VERSION BH842853.1 GI:21410104
    KEYWORDS GSS.
    SOURCE Trypanosoma cruzi.
    ORGANISM Trypanosoma cruzi.
    Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae;
    Trypanosoma; Schizotrypanum.
    1 (bases 1 to 662)
    Myler,P.J., Aggarwal,G., Fazelinia,G., Mack,D., Marty,A., Munden,H.
    , Nelson,S., Pentony,M., Rinta,J., Robertson,L., Seyler,A., Sis,E.E.,
    , Stuart,K., Vogt,C., Worthey,E., El-Sayed,N.M., Chedin,E. and
    Andersson,B.
    Trypanosoma cruzi CL-Brener TC3 BAC-end sequencing
    Unpublished (2001)
    Other GSSs: TC3-57A4.TP
    Contact: Peter Myler
    Seattle Biomedical Research Institute
    4 Nickerson Street, Seattle, WA 98109, USA
    Tel: 206 284 8846
    Fax: 206 284 0313
    Email: mylerp@sbri.org
    Clones are derived from the Trypanosoma cruzi CL-Brener BAC library
    TC3. For clone availability, please contact Dr. Bjorn Andersson at
    Uppsala University (bjorn.andersson@genpat.uu.se).
    Seq primer: 17
    Class: BAC ends.

FEATURES
    source
        location/Qualifiers
            1..662
                /organism="Trypanosoma cruzi"
                /strain="CL Brener"
                /db_xref="taxon:5693"
                /clone="TC3-57A4"
                /clone_id="TC3"
                /note="Vector: pBelosAC11, site_1: Hin dIII; Constructed
                for Uppsala University by Marie-Christine le Paslier in
                the Laboratory of Denis le Paslier at the Centre d'Etude
                du Polymorphisme Humain (CEPH), Paris, France. Briefly,
                Trypanosoma cruzi CL-Brener agarose embedded DNA (obtained
                from Dr. Franco de Silveira) was partially digested with
                Hin dIII. High molecular weight fragments were ligated in

```

pBelobAC11 digested with *Hin* dIII. The average insert size is 100 kb. Total clone coverage: approx. 33 X the haploid genome. "

```

Query Match      100.0%; Score 16; DB 17; length 662;
Best Local Similarity 100.0%; Pred. No. 4.7e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTTTGGCAATTTTC 16
          |||||||
Db       531 TTGTTGCAATTTTC 516

```

RESULT 7					
BE960519/c					
LOCUS	BE960519	926 bp	mRNA	linear	EST 04-OCT-2000
DEFINITION	60165332782 NIH_MGC_58 Homo sapiens CDNA clone IMAGE:3826193 3',				
ACCESSION	mRNA sequence.				
VERSION	BE960519				
KEYWORDS	BE960519.1	GI:10571224			
SOURCE	EST.				
	human.				

ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
homo sapiens					
Elkayam, Melazo					
Chordata: Vertebrata: Euteleostomi:					
Mammalia: Eutheria: Primates: Catarrhini: Hominoide: Homo.					
1 (bases 1 to 926)					
NIH-MGC http://mgc.ncl.nih.gov/ .					
National Institutes of Health, Mammalian Gene Collection (MGC)					
Unpublished (1999)					
Contact: Robert Strassberg, Ph.D.					

CIGNA Library Preparation: CLONETECH Laboratories, Inc.
 CIGNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA sequencing by: Invitae Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
 Plate: HICM493 ROW: F COLUMN: 18
 High quality sequence start: 22
 High quality sequence stop: 43.
 Location/Qualifiers
 1..926

/organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_image="3826193"
 /clone_1ib="NH.MGC.58"
 /tissue_type="hypertrophoma"
 /lab_host="DH10B (rt phage-resistant)"
 /note="Organ: kidney; Vector: pDMR-LIB (Clontech); site_1:
 Sfil (ggcgccgtcgcgc); Site_2: Sfil (ggccattatggcc);
 Double-stranded cDNA was prepared from cell line RNA. 5'
 and 3' adaptors were used in cloning as follows: 5'
 adaptor sequence: 5'-cagcagcagcagcagcagcagcagc-3' and 3' adaptor
 sequence: 5'-ttttctagaccgcagcagcagcagcagcagc-3' (30)BH-3'
 (where B = A, C, G, or G and N = A, C, G, or T). Average
 insert size 1.35 Kb (range 0.9-4.0 kb). 15/15 colonies
 contained inserts by PCR. This library was enriched for
 full-length clones and was constructed by Clontech
 Laboratories (Palo Alto, CA)."

```

Query Match      100.0%; Score 16; DB 12; length 926;
Best Local Similarity 100.0%; Pred. No. 4, 6e+03;
Matches 16; conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 TTGTTGTCACAATTTC 16
          |||#####
Db      464 TTGTTGTCACAATTTC 449

```

RESULT 8
BG948370

LOCUS	BG948370	132 bp	mRNA	linear	EST 12-JUN-2007
DEFINITION	IP1_12_H06.g1_A002	Immature panicle 1 (IP1) Sorghum bicolor cDNA, mRNA sequence.			
ACCESSION	BC048370				

ACCESSION	BG948370
VERSION	BG948370.1
KEYWORDS	EST.
SOURCE	sorghum.

ORGANISM	Sorghum bicolor
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

REFERENCE
1 (pages 1 to 132)

AUTHORS	Klein, R.K., Cordonnier-Pratt, M.-M., Gingie, A., Sudman, M. and P. L.H.
TITLE	An EST database from Sorghum: developing preanthesis panicles
JOURNAL	Unpublished (2001)
COMMENT	Contact: Cordonnier-Pratt MM

Laboratory for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpatt@uga.edu

Sequences have been trimmed to exclude PolyA₁ vector and regions below Phred quality 16. The threshold for high quality sequence is 20. Three-prime sequences, which are obtained with PolyMix or T7 sequencing primer, are presented as the reverse complement.

Seq primer: T7

High quality sequence start: 12
High quality sequence stop: 105
PolyA=yes.

FEATURES	Location/Qualifiers
source	1. .132

```

/organism="Sorghum bicolor"
/cultivar="BTx623"
/db_xref="taxon:4558"
/clone_id="Immature panicle 1 (1P1)"
/note="Organ: Developing preanthesis panicles; Vector:
pBluescript II SK(-) from lambda Zap II; Site: 1: XhoI;
Site: 2: EcoRI. The library was made from poly-A RNA in the
cloning vector lambda Zap II. Clones to be sequenced were
prepared by mass excision."
BASE COUNT      36 a      18 c      29 g      49 t
ORIGIN

```

```

QY      1  TTGGTGTGCATTTTT 15
          |||
Db      42 TTGTTGTCATTTTT 56

Query Match      93.8%; Score 15; DB 13; Length 13;
Best Local Similarity 100.0%; Pred. No. 1.3e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 9		
AM480273/c	LOCUS	200 bp mRNA linear EST 09-JUL-2006
	DEFINITION	30583 MRC iPig Sus scrofa cDNA 5', mRNA sequence.
	ACCESSION	AM480273
	VERSION	AM480273.1 GI:7050379
	KEYWORDS	
	SOURCE	EST.
	ORGANISM	Pig. (<i>Sus scrofa</i>)

ORGANISM	TITLE
<i>Sus scrofa</i>	
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Suidae; Suidae; Sus.	
1 (bases 1 to 200)	
Fahrenkrug, S.C., Feking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E., Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W. and Keele, J.W.	
	Design and use of two pooled tissue normalized cDNA libraries for

JOURNAL
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4356
Fax: 402 762 4390
Email: smith@email.marc.usda.gov

Single pass sequencing. Bases called and trimmed with phred v0.980904.e. Vector identified by cross_match with the -minscore 20 and -minmatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTAGACCAT
BACKWARD: GTTTCCTCAGTCAGACG
Plate: 9 row: B column: 18
Seq primer: ATTACGCTGACACTATAG.
Location/Qualifiers
1. 200
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_id="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: NotI; Site_2: SalI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."

BASE COUNT 89 a 36 c 30 g 45 t
ORIGIN

Query Match 93.8%; Score 15; DB 10; Length 200;
Best Local Similarity 100.0%; Pred. No. 1.3e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTTGCAATTTT 15
|||||
DB 106 TTGTTGCAATTTT 92

RESULT 10
AF049574/c 218 bp mRNA linear EST 17-APR-1998
LOCUS AF049574 Human activated dendritic cell mRNA Homo sapiens cDNA
DEFINITION clone TA03, mRNA sequence.
ACCESSION AF049574
VERSION AF049574.1 GI:3046958
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 218)
AUTHORS Zhao, Z., Huang, X., Cao, X. and Li, N.
TITLE mRNA Differential Display of dendritic cell by amplified fragment length polymorphism
JOURNAL Unpublished (1998)
COMMENT Contact: Zhao, Zhongliang
Department of Molecular Immunology
Shanghai Brilliance Biotechnology
No. 800, Xiangyin Road, Shanghai, SH 200433, P.R. China
Email: xtao@public.sta.net.cn

FEATURES
SOURCE

1. 218
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="TA03"
/clone_id="Human activated dendritic cell mRNA"
/cell_type="activated dendritic"
/note="Organ: soft tissue; Vector: pT73D-Pac (Pharmacia)
with a modified polylinker; Site_1: Not I; Site_2: Eco RI;
1st strand cDNA was primed with a Not I - oligo(dT) primer
[5'-AACGTGACAGATTCGGCGCGAATCGATGATTTTCTTTT-3']
double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I

BASE COUNT 79 a 46 c 31 g 58 t 4 others
ORIGIN

Query Match 93.8%; Score 15; DB 9; Length 218;
Best Local Similarity 100.0%; Pred. No. 1.3e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTTGCAATTTT 15
|||||
DB 166 TTGTTGCAATTTT 152

RESULT 11
LOCUS B0556095
DEFINITION B0556095 251 bp mRNA linear EST 20-JUN-2002
H4038A08 3' mRNA sequence.
ACCESSION B0556095
VERSION B0556095.1 GI:21456983
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 251)
AUTHORS Vanburen, V., Piao, Y., Dudekula, D. B., Qian, Y., Carter, M. G., Martin
P. R., Stagg, C. A., Bassey, V., Alpa, K., Hamatani, T., Kargul, G. J.,
Luo, A. G. and Ko, M. S. H.
Assembly, verification, and initial annotation of NIA 7.4K mouse
cDNA clone set
Unpublished (2002)
Contact: Yong Qian
Laboratory of Genetics
National Institute on Aging/National Institutes of Health
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA
Email: cdna@igsn.gcr.nia.nih.gov
This clone set has been freely distributed to the community. Please
visit http://19sun.grc.nia.nih.gov/cDNA/NIA_7_4K.html for details.
Plate: H4038 row: A column: 08
Seq primer: -21M13 Forward
High quality sequence stop: 251
POLYA=Yes.

FEATURES
SOURCE

1. 251
/organism="Mus musculus"
/strain="C57BL/6"
/db_xref="taxon:H4038A08-3"
/db_xref="taxon:10090"
/clone="H4038A08"
/clone_id="NIA Mouse 7.4K cDNA Clone Set"
/sex="mixed"
/dev_stage="mixed"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI. This
clone is among a rearranged set of 7,407 clones from more
than 20 cDNA libraries."

BASE COUNT 66 a 51 c 46 g 88 t
ORIGIN

Query Match 93.8%; Score 15; DB 14; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e+04;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 TTGTTGCAATTTTTC 16
|||||
DB 226 TTGTTGCAATTTTTC 240

RESULT 12
AL731053/c

LOCUS AL731053 267 bp mRNA linear EST 18-APR-2002
 DEFINITION AL731053 Danio rerio embryonic inner ear subtracted cDNA Danio
 rerio cDNA clone BND0A120ZG09 5', mRNA sequence.
 ACCESSION AL731053
 VERSION AL731053.1 GI:20195657
 KEYWORDS EST.
 SOURCE zebrafish.
 ORGANISM Danio rerio
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes
 ; Cyprinidae; Danio.
 REFERENCE 1 (bases 1 to 267)
 AUTHORS Coimbra,R., Weil,D., Brottier,P., Blanchard,S., Levi,M., Hargelin
 ,J.P., Weissenbach,J. and Petit,C.
 A subtracted cDNA library from the zebrafish (Danio rerio)
 embryonic inner ear
 TITLE embryonic inner ear
 JOURNAL Unpublished (2002)
 COMMENT Contact: Genoscope
 Genoscope - Centre National de Sequencage
 BP 191 91006 EVRY cedex - France
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
 FEATURES
 source location/Qualifiers
 1.267
 /organism="Danio rerio"
 /db_xref="taxon:7955"
 /clone="BND0A120ZG09"
 /clone_1lb="Danio rerio embryonic inner ear subtracted
 cDNA"
 /tissue_type="inner ear"
 /dev_stage="embryonic"
 /note="subtracted cDNA library"
 BASE COUNT 72 a 55 c 78 g 62 t
 ORIGIN
 Query Match 93.8%; Score 15; DB 9; Length 267;
 Best local Similarity 100.0%; Pred. No. 1.3e+04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 2 TGTGTGCATTTTC 16
 |||||||||||||
 Db 249 TGTGTGCATTTTC 235
 RESULT 13 267 bp mRNA linear EST 27-JUN-2000
 LOCUS B0113077
 DEFINITION B0113077 RIKEN full-length enriched, adult male urinary bladder Mus
 musculus cDNA clone 9530038H16 3' similar to D30782 Mouse mRNA for
 epiregulin, mRNA sequence.
 ACCESSION B0113077
 VERSION B0113077.1 GI:8765645
 KEYWORDS EST.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 267)
 AUTHORS Kono,H., Aizawa,K., Akahira,S., Akiyama,Y., Arakawa,T., Carninci
 ,P., Endo,T., Fukuda,S., Fukunishi,Y., Hara,A., Hayatsu,N.,
 Hirozane,T., Hori,F., Ishii,Y., Ishikawa,J., Ishikawa,T., Itoh,M.,
 Izawa,M., Kadota,K., Kagawa,I., Kai,C., Kawai,J., Kikuchi,N.,
 Kiyosawa,H., Kojima,Y., Kondo,S., Koya,S., Kurihara,C., Kusakabe,M.,
 Matsuyama,T., Miki,R., Mizuno,Y., Nakamura,M., Oda,H., Okazaki,Y.,
 Ono,T., Owa,C., Saito,H., Sakai,C., Sato,K., Shibata,K., Shibata,Y.,
 Shigemoto,Y., Shinagawa,A., Shikata,T., Sogabe,Y., Sugahara,Y.,
 Suzuki,H., Suzuki,H., Tagawa,A., Takahashi,F., Tomimaga,N., Toyota,
 ,T., Tsunoda,Y., Watabiki,A., Watanabe,S., Yamamura,T., Yamanaoka,I.,
 ,Yano,R., Yasunishi,A., Yokota,T., Yoshida,K., Yoshiki,A., Yoshino
 ,M., Muramatsu,M. and Hayashizaki,Y.
 RIKEN Mouse ESTs (Kono,H., et al.)
 TITLE Unpublished (2000)
 JOURNAL Laboratory for Genome Exploration Research Group, RIKEN Genomic

Sciences Center(GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Saito-cho, Tsukuba-shi, Ibaraki, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-res@sc.riken.go.jp,
 URL:http://genome.gsc.riken.go.jp/
 Carninci,P., Nishiyama,Y., Westover,A., Itoh,M., Nagakawa,S., Sasaki
 N., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
 Thermosensitization and thermoregulation of the nucleobase enzymes by
 trehalose and its application for the synthesis of full length
 cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)
 Itoh,M., Katsunari,T., Akiyama,Y., Shibata,K., Izawa,M., Kawai,J.,
 Tomaru,Y., Carninci,P., Shibata,Y., Ozawa,Y., Muramatsu,M., Okazaki
 ,Y. and Hayashizaki,Y.
 Automated filtration-based high-throughput plasmid preparation
 system. Genome Res. 9 (5), 463-470 (1999)
 Carninci,P. and Hayashizaki,Y.
 High-efficiency full-length cDNA cloning. Methods Enzymol. 303,
 19-44 (1999)
 Please visit our web site (http://genome.rtc.riken.go.jp) for
 further details.
 FEATURES
 source location/Qualifiers
 1.267
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /clone="9530038H16"
 /clone_1lb="RIKEN full-length enriched, adult male urinary
 bladder"
 /sex="male"
 /tissue_type="urinary bladder"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="Site_1: Saito; Site_2: BamHI; cDNA library was
 prepared and sequenced in Mouse Genome Encyclopedia
 Project of Genome Exploration Research Group in Riken
 Genomic Sciences Center and Genome Science Laboratory in
 RIKEN. Division of Experimental Animal Research in Riken
 contributed to prepare mouse tissues. 1st strand cDNA was
 primed with a primer [5'
 GAGAGAGAGATTCAGAGATTCATTTTCTTTTCTTTTCTTTT 3'], cDNA was
 prepared by using trehalose thermo-activated reverse
 transcriptase and subsequently enriched for full-length by
 cap-trapper. cDNA went through one round of normalization
 to Rot = 20.0 and subtraction to Rot = 370.0. Second
 strand cDNA was prepared with the primer adapter of
 sequence [5' GAGAGAGATTCAGAGATTCATTTTCTTTTCTTTTCTTTT
 3']. cDNA was cleaved with XhoI and BamHI. Vector: a
 modified plasmid pUC19 KS(+) after bulk excision from
 lambda FliC I."]
 BASE COUNT 79 a 70 c 30 g 88 t
 ORIGIN
 Query Match 93.8%; Score 15; DB 10; Length 267;
 Best local Similarity 100.0%; Pred. No. 1.3e+04;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTGTGCATTTTC 15
 |||||||||||||
 Db 192 TTGTGTGCATTTTC 206
 RESULT 14 267 bp mRNA linear EST 15-JUN-2001
 LOCUS B1054222
 DEFINITION B1054222 PM3-GN0375-310101-003-f08 GN0375 Homo sapiens cDNA, mRNA sequence.
 ACCESSION B1054222
 VERSION B1054222.1 GI:14461752
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 18:49:02 ; Search time 29.7143 Seconds
(without alignments)
241.912 Million cell updates/sec

Title: US-09-803-736-1022_COPY_10615_10630
Perfect score: 16
Sequence: 1 tttgtgtcaatttttc 16

Scoring table: IDENTITY_MTC
Gapop 10.0, Gapext 1.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCU_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCUUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	15	93.8	499	9	US-09-908-931B-51	Sequence 51, Appl
2	15	93.8	2000	9	US-09-938-842A-4380	Sequence 4380, Ap
3	14.4	90.0	470	10	US-09-960-352-5575	Sequence 5575, Ap
4	14.4	90.0	907	10	US-09-939-980-161	Sequence 161, App
5	14.4	90.0	1371	10	US-09-815-242-7207	Sequence 7207, Ap
6	14.4	90.0	2000	9	US-09-938-842A-3393	Sequence 3393, Ap
7	14.4	90.0	2793	9	US-09-938-842A-394	Sequence 394, App
8	14.4	90.0	3199	9	US-09-945-249-10	Sequence 10, Appl
9	14.4	90.0	5840	10	US-09-070-927A-31	Sequence 31, Appl
10	14.4	90.0	6200	10	US-09-764-850-1185	Sequence 1185, Ap
11	14.4	90.0	7045	10	US-09-764-850-1185	Sequence 3794, Ap
12	14.4	90.0	19820	10	US-09-764-850-1185	Sequence 861, App
13	14.4	90.0	21721	10	US-09-764-850-1185	Sequence 116, App
14	14.4	90.0	41936	10	US-09-764-850-1185	Sequence 3, Appl1
15	14.4	90.0	202001	10	US-09-734-674-3	Sequence 3, Appl1
16	14.4	90.0	203654	10	US-09-820-905-3	Sequence 3, Appl1
17	14.4	90.0	326014	10	US-09-731-231A-3	Sequence 643, App
18	14	87.5	401	9	US-09-946-807-643	Sequence 1434, Ap
19	14	87.5	401	9	US-09-946-807-1434	

C 20	14	87.5	401	10	US-09-795-668-643	Sequence 643, App
C 21	14	87.5	401	10	US-09-795-668-1434	Sequence 1434, App
C 22	14	87.5	401	10	US-09-795-668-643	Sequence 643, App
C 23	14	87.5	401	10	US-09-795-668-1434	Sequence 1434, App
C 24	14	87.5	402	9	US-09-946-807-1433	Sequence 1433, App
C 25	14	87.5	402	9	US-09-946-807-1433	Sequence 1433, App
C 26	14	87.5	402	10	US-09-795-668-1433	Sequence 1433, App
C 27	14	87.5	402	10	US-09-795-668-1433	Sequence 1433, App
C 28	14	87.5	402	10	US-09-795-668-1433	Sequence 1433, App
C 29	14	87.5	402	10	US-09-795-668-1433	Sequence 1433, App
C 30	14	87.5	573	10	US-09-864-761-8664	Sequence 8664, App
C 31	14	87.5	787	10	US-09-765-272-97	Sequence 97, Appl
C 32	14	87.5	811	10	US-09-765-272-205	Sequence 205, App
C 33	14	87.5	868	9	US-09-938-842A-5290	Sequence 5290, App
C 34	14	87.5	913	10	US-09-834-975-274	Sequence 274, App
C 35	14	87.5	2000	9	US-09-938-842A-2770	Sequence 2770, App
C 36	14	87.5	2000	9	US-09-938-842A-3221	Sequence 3221, App
C 37	14	87.5	2613	10	US-09-822-830A-518	Sequence 518, App
C 38	14	87.5	3106	8	US-08-837-459-21	Sequence 21, Appl
C 39	14	87.5	3131	8	US-08-837-459-20	Sequence 20, Appl
C 40	14	87.5	32195	10	US-09-764-869-2017	Sequence 2017, App
C 41	14	87.5	32219	10	US-09-764-869-2016	Sequence 2016, App
C 42	13.4	83.8	90	9	US-10-016-634A-38	Sequence 38, Appl
C 43	13.4	83.8	117	10	US-09-864-761-26198	Sequence 26198, A
C 44	13.4	83.8	124	10	US-09-969-373-1038	Sequence 1038, App
C 45	13.4	83.8	125	10	US-09-969-373-1036	Sequence 1036, App

ALIGNMENTS

RESULT 1
US-09-908-931B-51
Sequence 51, Application US/09908931B
Patent No. US20020164756A1
GENERAL INFORMATION:
APPLICANT: Pearson, James D.
APPLICANT: Slightom, Jerry
APPLICANT: Chosay, John G.
APPLICANT: Shindhaber, Dean L.
TITLE OF INVENTION: Complete Nucleotide Sequence of Staphylococcus aureus
TITLE OF INVENTION: Ribosomal Protein Gene, S16 and Methods for the
FILE REFERENCE: S16 ribosomal protein
CURRENT APPLICATION NUMBER: US/09/908, 931B
CURRENT FILING DATE: 2001-07-19
NUMBER OF SEQ ID NOS: 58
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 51
LENGTH: 499
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-908-931B-51

Query Match 93.8%; Score 15; DB 9; Length 499;
Best Local Similarity 100.0%; Pred. No. 2.2e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Ox 1 TTTGTGCAATTTT 15
DB 395 TTTGTGCAATTTT 409

RESULT 2
US-09-938-842A-4380
Sequence 4380, Application US/09938842A
Patent No. US20020160378A1
GENERAL INFORMATION:
APPLICANT: Harper, Jeff
APPLICANT: Krepes, Joel
APPLICANT: Wang, Xun
APPLICANT: Zhu, Tong
TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAIN

```

; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRIPI300-3
; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 4380
; LENGTH: 2000
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-4380

Query Match          93.8%; Score 15; DB 9; Length 2000;
Best Local Similarity 100.0%; Pred. No. 2.4e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTT 15
DB 1159 TTGTTGCAATTTT 1173

RESULT 3
US-09-960-352-5575/c
; Sequence 5575, Application US/09960352
; Patent No. US2002017139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Hyatt, John C.
; APPLICANT: Mathalagan, Nageppan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 5575
; LENGTH: 470
; TYPE: DNA
; ORGANISM: Bos taurus
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (451)
; OTHER INFORMATION: unsure at all n locations
; OTHER INFORMATION: Clone ID: 24-LIB3058-021-Q1-K1-F7
US-09-960-352-5575

Query Match          90.0%; Score 14.4; DB 10; Length 470;
Best Local Similarity 93.8%; Pred. No. 4.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTT 16
DB 240 TTATGTCAATTTT 225

RESULT 4
US-09-939-980-161
; Sequence 161, Application US/09939980
; Patent No. US2002008234A1
; GENERAL INFORMATION:
; APPLICANT: Black, Michael
; APPLICANT: Burnham, Martin
; APPLICANT: Hodgson, John
; APPLICANT: Knowles, David
; APPLICANT: Lometto, Michael
; APPLICANT: Nicholas, Richard
; APPLICANT: Pratt, Julie

; TITLE OF INVENTION: No. US2002008234A1el Prokaryotic Polynucleotides,
; Polypeptides and their uses
; NUMBER OF SEQUENCES: 534
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Smithkline Beecham Corporation
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-0939
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/939,980
; FILING DATE: 27-Aug-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/936,165
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Gilmul, Edward R
; REGISTRATION NUMBER: 38,891
; REFERENCE/DOCKET NUMBER: P50549
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 610-270-4478
; TELEFAX: 610-270-5090
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 161:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 907 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; SEQUENCE DESCRIPTION: SEQ ID NO: 161:
US-09-939-980-161

Query Match          90.0%; Score 14.4; DB 10; Length 907;
Best Local Similarity 93.8%; Pred. No. 4.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTT 16
DB 862 TTGTTGTAATTTT 877

RESULT 5
US-09-815-242-7207/c
; Sequence 7207, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
```

RESULT 9
US-09-070-927A-31
; Sequence 31, Application US/09070927A
; Patent No. US20020120116A1
; GENERAL INFORMATION:

```
APPLICANT: Charles A. Kunsch
           Patrick J. Dillon
           Steven Barash
TITLE OF INVENTION: Enterococcus faecialis Polynucleotides and Polypeptides
NUMBER OF SEQUENCES: 962
CORRESPONDENCE ADDRESS:
ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/070,927A
FILING DATE: 04-May-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/046,655
FILING DATE: 1997-05-16
APPLICATION NUMBER: 60/044,031
FILING DATE: 1997-05-06
APPLICATION NUMBER: 60/066,009
FILING DATE: 1997-11-14
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB369
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 31:
SEQUENCE CHARACTERISTICS:
LENGTH: 5840 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 31:
US-09-070-927A-31
Query Match          90.0%; Score 14.4; DB 10; Length 5840;
Best Local Similarity 93.8%; Pred. No. 4.9e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGTGTCAATTTTC 16
Db 4571 TTTTGTCAATTTTC 4586
RESULT 10
US-09-764-860-1185
Sequence 1185, Application US/09764860
Patent No. US20020094953A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC008
CURRENT APPLICATION NUMBER: US/09/764,860
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PAM or file wrapper
NUMBER OF SEQ ID NOS: 1198
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1185
LENGTH: 6209
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-860-1185
Query Match          90.0%; Score 14.4; DB 10; Length 6209;
```

```
Best Local Similarity 93.8%; Pred. No. 4.9e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGTGTCAATTTTC 16
Db 4997 TTGTGTCAATTTTC 5012
RESULT 11
US-09-764-877-3794/c
Sequence 3794, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PAM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 3794
LENGTH: 7045
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-877-3794
Query Match          90.0%; Score 14.4; DB 10; Length 7045;
Best Local Similarity 93.8%; Pred. No. 4.9e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGTGTCAATTTTC 16
Db 5272 TTTTGTCAATTTTC 5257
RESULT 12
US-09-764-877-2713
Sequence 2713, Application US/09764877
Patent No. US20020147140A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PC005
CURRENT APPLICATION NUMBER: US/09/764,877
CURRENT FILING DATE: 2001-01-17
Prior application data removed - refer to PAM or file wrapper
NUMBER OF SEQ ID NOS: 4031
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2713
LENGTH: 19820
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (13245)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (13259)
OTHER INFORMATION: n equals a,t,g, or c
US-09-764-877-2713
Query Match          90.0%; Score 14.4; DB 10; Length 19820;
Best Local Similarity 93.8%; Pred. No. 5.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGTGTCAATTTTC 16
Db 6760 TTGTGTCAATTTTC 6775
RESULT 13
US-09-764-853-861/c
```



```

; Sequence 861, Application US/09764853
; Patent No. US2002090672A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P1206
; CURRENT APPLICATION NUMBER: US/09/764,853
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 939
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 861
; LENGTH: 21721
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-853-861

Query Match          90.0%; Score 14.4; DB 10; Length 21721;
Best Local Similarity 93.8%; Pred. No. 5.2e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTGTTGTCATTTTTC 16
        ||||| ||||| |||||
Db 11928 TTGTTGTCATTTTTC 11913

RESULT 14
US-09-967-768A-116/c
; Sequence 116, Application US/09967768A
; Patent No. US2002015087A1
; GENERAL INFORMATION:
; APPLICANT: Augustus, Meena
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Signatu
; FILE REFERENCE: 689290-72
; CURRENT APPLICATION NUMBER: US/09/967,768A
; CURRENT FILING DATE: 2001-09-28
; PRIOR APPLICATION NUMBER: US/60/236,109
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,034
; PRIOR FILING DATE: 2000-09-28
; PRIOR APPLICATION NUMBER: US/60/236,111
; PRIOR FILING DATE: 2000-09-28
; NUMBER OF SEQ ID NOS: 325
; SOFTWARE: Patent Ver. 3.0
; SEQ ID NO 116
; LENGTH: 41936
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-967-768A-116

Query Match          90.0%; Score 14.4; DB 10; Length 41936;
Best Local Similarity 93.8%; Pred. No. 5.4e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTGTTGTCATTTTTC 16
        ||||| ||||| |||||
Db 28725 TTGTTGTCATTTTTC 28710

RESULT 15
US-09-734-674-3/c
; Sequence 3, Application US/09734674
; Patent No. US2002081648A1
; GENERAL INFORMATION:
; APPLICANT: Wei, Ming-Hui et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: C1001018
; CURRENT APPLICATION NUMBER: US/09/734,674
; CURRENT FILING DATE: 2000-12-13
; NUMBER OF SEQ ID NOS: 4

```

```

; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 202001
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: (1)...(202001)
; OTHER INFORMATION: n = A,T,C or G
US-09-734-674-3

Query Match          90.0%; Score 14.4; DB 10; Length 202001;
Best Local Similarity 93.8%; Pred. No. 5.3e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      1 TTGTTGTCATTTTTC 16
        ||||| ||||| |||||
Db 75167 TTGTTGTCATTTTTC 75152

Search completed: January 23, 2003, 20:35:05
Job time : 55.7143 secs

```


GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 19:22:47; Search time 28.3429 Seconds

(without alignments)
173.124 Million cell updates/sec

Title: US-09-803-736-1022_COPY_10615_10630

Perfect score: 16

Sequence: 1 ttgtgtcatttttc 16

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2_6/p10data/2/ina/5A.COMB.seq:*
2: /cgn2_6/p10data/2/ina/5B.COMB.seq:*
3: /cgn2_6/p10data/2/ina/6A.COMB.seq:*
4: /cgn2_6/p10data/2/ina/6B.COMB.seq:*
5: /cgn2_6/p10data/2/ina/PCNUS.COMB.seq:*
6: /cgn2_6/p10data/2/ina/backfilee1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	15	93.8	1974	1	US-08-413-135-3
2	15	93.8	1974	4	US-08-971-395-3
3	14.4	90.0	907	4	US-08-936-165A-161
4	14.4	90.0	1201	1	US-08-181-271A-26
5	14.4	90.0	1201	1	US-08-449-315-26
6	14.4	90.0	1201	1	US-08-444-803-26
7	14.4	90.0	1201	1	US-08-449-043-26
8	14.4	90.0	1201	1	US-08-456-265A-26
9	14.4	90.0	1201	1	US-08-455-416-26
10	14.4	90.0	1201	1	US-08-455-244-26
11	14.4	90.0	1201	1	US-08-454-876-26
12	14.4	90.0	1201	2	US-08-457-364-26
13	14.4	90.0	1201	2	US-08-456-262-26
14	14.4	90.0	1201	2	US-08-456-240-26
15	14.4	90.0	1201	2	US-08-455-736-26
16	14.4	90.0	1201	2	US-08-971-217-26
17	14.4	90.0	1201	4	US-09-350-600-26
18	14.4	90.0	3198	4	US-08-842-306B-48
19	14.4	90.0	3198	4	US-08-838-973B-48
20	14.4	90.0	3464	4	US-09-318-448-30
21	14.4	90.0	6201	2	US-08-790-912-1
22	14	87.5	787	3	US-08-961-083-97
23	14	87.5	811	3	US-08-961-083-205
24	14	87.5	1959	4	US-09-201-641-1
25	14	87.5	3106	4	US-08-840-466A-21
26	14	87.5	3106	4	US-09-696-188B-21
27	14	87.5	3131	4	US-08-840-466A-20

C	28	14	87.5	3131	4	US-09-696-188B-20	Sequence 20, Appl
	29	14	87.5	3178	3	US-08-968-563-7	Sequence 7, Appl
	30	14	87.5	3178	3	US-08-969-683A-7	Sequence 7, Appl
	31	14	87.5	3178	4	US-09-297-928-3	Sequence 3, Appl
C	32	14	87.5	12127	4	US-08-961-527-148	Sequence 148, App
	33	14	87.5	43360	4	US-09-453-702B-206	Sequence 206, App
	34	14	87.5	45325	4	US-09-453-702B-261	Sequence 261, App
	35	13.4	83.8	21	1	US-08-753-147-123	Sequence 123, App
	36	13.4	83.8	72	1	US-08-413-110-122	Sequence 122, App
	37	13.4	83.8	72	3	US-08-473-446-122	Sequence 122, App
C	38	13.4	83.8	448	4	US-09-149-476-234	Sequence 234, App
	39	13.4	83.8	508	4	US-09-149-476-69	Sequence 69, Appl
C	40	13.4	83.8	518	4	US-09-280-116-245	Sequence 245, App
C	41	13.4	83.8	699	4	US-08-476-102A-6	Sequence 6, Appl
C	42	13.4	83.8	715	4	US-08-998-416-788	Sequence 788, Appl
	43	13.4	83.8	749	1	US-08-579-667-9	Sequence 9, Appl
	44	13.4	83.8	764	4	US-08-858-207A-224	Sequence 224, App
	45	13.4	83.8	786	5	PCT-US95-08295-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-413-135-3
Sequence 3, Application US/08413135
Patent No. 5689042
GENERAL INFORMATION:
APPLICANT: Amastio, Richard M
TITLE OF INVENTION: Transgenic Plants with Altered
TITLE OF INVENTION: Senseless Characteristics
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESSES:
ADDRESSER: Quarles & Brady
STREET: 1 South Plinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53703
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/413,135
FILING DATE:
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J
REGISTRATION NUMBER: 27,386
REFERENCE/DOCKET NUMBER: 960296.92808
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-251-5000
TELEFAX: 608-251-9166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1974 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "SAG13 Promoter DNA"
US-08-413-135-3
Query Match 93.8%; Score 15; DB 1; Length 1974;
Best Local Similarity 100.0%; Pred. No. 93;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 1115 TTGTTGCAATTTT 1129
OY 1 TTGTTGCAATTTT 15
|||||
|||||

RESULT 2

US-08-971-395-3

; Sequence 3, Application US/08971395

; Patent No. 6359197

; GENERAL INFORMATION:

; APPLICANT: Amasino, Richard M

; APPLICANT: No. 6359197, Too-Sun

; APPLICANT: Gan, Sheng

; TITLE OF INVENTION: Transgenic Plants with Altered

; TITLE OF INVENTION: Senseless Characteristics

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Quarles & Brady

; STREET: 1 South Pinckney Street

; CITY: Madison

; STATE: WI

; COUNTRY: US

; ZIP: 53701-2113

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/971,395

; FILING DATE:

; CLASSIFICATION: 800

; ATTORNEY/AGENT INFORMATION:

; NAME: Seay, Nicholas J.

; REGISTRATION NUMBER: 27386

; REFERENCE/DOCKET NUMBER: 960296.94908

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 608-251-9166

; TELEFAX: 608-251-5000

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1974 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; US-08-971-395-3

Query Match 93.8%; Score 15; DB 4; Length 1974;

Best Local Similarity 100.0%; Pred. No. 93;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTT 15

DB 1115 TTGTTGCAATTTT 1129

RESULT 3

US-08-936-165A-161

; Sequence 161, Application US/08936165A

; Patent No. 6348582

; GENERAL INFORMATION:

; APPLICANT: Black, Michael

; APPLICANT: Burnham, Martin

; APPLICANT: Hodgson, John

; APPLICANT: Knowles, David

; APPLICANT: Lomotto, Michael

; APPLICANT: Nicholas, Richard

; APPLICANT: Pratt, Julie

; APPLICANT: Reichard, Richard

; APPLICANT: Rosenberg, Martin

; APPLICANT: Ward, Judith

; TITLE OF INVENTION: Polypeptides and Their Uses

; TITLE OF INVENTION: Polypeptides and Their Uses

; NUMBER OF SEQUENCES: 534

; CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation

STREET: 709 Swedeland Road

CITY: King of Prussia

STATE: PA

COUNTRY: USA

ZIP: 19406-0939

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/936,165A

FILING DATE: 24-SEP-1997

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,032

FILING DATE: 24-SEP-1996

ATTORNEY/AGENT INFORMATION:

NAME: Gimm, Edward R

REGISTRATION NUMBER: 38,891

REFERENCE/DOCKET NUMBER: P50549

TELECOMMUNICATION INFORMATION:

TELEPHONE: 610-270-4478

TELEFAX: 610-270-5090

TELEX:

INFORMATION FOR SEQ ID NO: 161:

SEQUENCE CHARACTERISTICS:

LENGTH: 907 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: Genomic DNA

; US-08-936-165A-161

Query Match 90.0%; Score 14.4; DB 4; Length 907;

Best Local Similarity 93.8%; Pred. No. 1.7e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTTC 16

DB 862 TTGTTGCAATTTTC 877

RESULT 4

US-08-181-271A-26/C

; Sequence 26, Application US/08181271A

; Patent No. 5614395

; GENERAL INFORMATION:

; APPLICANT: Ryals, John A.

; APPLICANT: Alexander, Danny C.

; APPLICANT: Beck, James J.

; APPLICANT: Duesing, John H.

; APPLICANT: Friedlich, Leslie B.

; APPLICANT: Goodman, Robert M.

; APPLICANT: Harms, Christian

; APPLICANT: Meins, Jr., Frederick

; APPLICANT: Montoya, Alice

; APPLICANT: Moyer, Mary B.

; APPLICANT: Neuhaus, Jean-Marc

; APPLICANT: Payne, George B.

; APPLICANT: Sperison, Christoph

; APPLICANT: Stinson, Jeffrey R.

; APPLICANT: Uknes, Scott J.

; APPLICANT: Ward, Eric R.

; APPLICANT: Williams, Shericea C.

; TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC

; TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF

; NUMBER OF SEQUENCES: 106

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: CIBA-GEIGY Corporation

; STREET: 7 Skyline Drive

; CITY: Hawthorne

STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/181,271A
FILING DATE: 13-JAN-94
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/769,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rimer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19625/P1/CSC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-181-271A-26

Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Q7 1 TTGTTGCAATTTTC 16
Db 152 TTGTTGCAATTTTC 137

RESULT 5
US-08-449-315-26/c
Sequence 26, Application US/08449315
Patent No. 5650505
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melns, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, SheriCCA C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY CORPORATION
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,315
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Rimer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PI/CSC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-449-315-26

Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGTCAATTTTC 16
DB 152 TTGTGTCAATTTTC 137

RESULT 6
US-08-444-803-26/C
Sequence 26, Application US/08444803
Patent No. 5654414
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Weins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Meyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Ukens, Scott J.
APPLICANT: Ward, Eric R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSER: CTBA-GRIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York

COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/444,803
FILING DATE: 19-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PI/CSC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ. ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-444-803-26

Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;

Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTTGCAATTTTC 16
||||| |||||
Db 152 TTGTTGCAATTTTC 137

RESULT 7
US-08-449-043-26/c
Sequence 26, Application US/08449043
Patent No. 5689044
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesling, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean Marc
APPLICANT: Payne, George B.
APPLICANT: Speetison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/449,043
FILING DATE: 24-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504

FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/C/GC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-449-043-26
MOLECULE TYPE: CDNA
Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 TTGTTGCAATTTTC 16
||||| |||||
Db 152 TTGTTGCAATTTTC 137

RESULT 8
US-08-456-265A-26/c
Sequence 26, Application US/08456265A
Patent No. 5767369
GENERAL INFORMATION:
APPLICANT: Alexander, Danny C.
APPLICANT: Ryals, John A.
APPLICANT: Goodman, Robert M.
APPLICANT: Stinson, Jeffrey R.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 111
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 520 White Plains Road, P.O. Box 2005
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,265A
FILING DATE: 31-MAY-95
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/181,271
FILING DATE: 13-JAN-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Melus, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: S-19825/PI/CGC 1727/DIV10
TELEPHONE: (919)541-8587
TELEFAX: (919)541-8589
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-456-265A-26

Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTTCATTTTC 16
||||| |||||||
DB 152 TTGTTGCCATTTTC 137

RESULT 9
; Sequence 26, Application US/08455416
; Patent No. 5777200

GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Melus, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhans, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Udes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Sheriella C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455,416
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIORITY APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1201 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 OS-08-435-416-26

Query Match	90.0%;	Score 14,4;	DB 1;	Length 1201;
Best Local Similarity	93.8%;	Pred. No. 1.7e+02;		
Matches 15; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

QY	1	TTGTTGTCATTTTTC	16
Db	152	TTGTTGGCAATTTTTC	137

RESULT 10
US-08-455-244-26/c
; Sequence 26, Application US/08455244

GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedlich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Means, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Speilson, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Sherica C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/08/455, 244
FILING DATE: 31-MAY-1995

```

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
type: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA

```

```

Query Match      90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0

QY      1 TTGTTTCATTTTC 16
          |||||  |||||
Db       152 TTGTTGGCAATTTTC 137

RESULT 11
US-08-454-876-26/c
: Sequence 26, Application US/08454876

```

Patent No. 5804693
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/454,876
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/348,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CCG 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-454-876-26
Query Match 90.0%; Score 14.4; DB 1; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 TTGTGTCATTTTTC 16
DB 152 TTGTTGCATTTTTC 137
RESULT 12
US-08-457-364-26/C
Sequence 26, Application US/08457364
Patent No. 5847258
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Utnes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericca C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/457,364

FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-457-364-26

Query Match 90.0%; Score 14.4; DB 2; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Caps 0;

Oy 1 TTGTTGCAATTTTC 16
||||| |||||||
Db 152 TTGTTGCAATTTTC 137

RESULT 13
US-08-456-262-26/c

Sequence 26, Application US/08456262
Patent No. 5851766
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhaus, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Uknes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/456,262
FILING DATE: 31-MAY-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431

FILED DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CSC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-456-262-26

Query Match 90.0%; Score 14.4; DB 2; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGCAATTTTC 16
||||| |||||||
Db 152 TTGTGCAATTTTC 137

RESULT 14
US-08-456-240-26/C
Sequence 26, Application US/08456240
Patent No. 5856154
GENERAL INFORMATION:
APPLICANT: Ryals, John A.
APPLICANT: Alexander, Danny C.
APPLICANT: Beck, James J.
APPLICANT: Duesing, John H.
APPLICANT: Friedrich, Leslie B.
APPLICANT: Goodman, Robert M.
APPLICANT: Harms, Christian
APPLICANT: Meins, Jr., Frederick
APPLICANT: Montoya, Alice
APPLICANT: Moyer, Mary B.
APPLICANT: Neuhau, Jean-Marc
APPLICANT: Payne, George B.
APPLICANT: Sperison, Christoph
APPLICANT: Stinson, Jeffrey R.
APPLICANT: Ukes, Scott J.
APPLICANT: Ward, Eric R.
APPLICANT: Williams, Shericea C.
TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: CIBA-GEIGY Corporation
STREET: 7 Skyline Drive
CITY: Hawthorne
STATE: New York
COUNTRY: USA
ZIP: 10532
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/456,240
FILING DATE: 31-MAY-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/181,271
FILING DATE: 13-JAN-94
APPLICATION NUMBER: US 08/093,301
FILING DATE: 16-JUL-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/937,197
FILING DATE: 6-NOV-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/678,378
FILING DATE: 1-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/305,566
FILING DATE: 6-FEB-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/165,667
FILING DATE: 8-MAR-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/042,847
FILING DATE: 6-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/632,441
FILING DATE: 21-DEC-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/425,504
FILING DATE: 20-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,506
FILING DATE: 6-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/768,122
FILING DATE: 27-SEP-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/580,431
FILING DATE: 7-SEP-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/368,672
FILING DATE: 20-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/329,018
FILING DATE: 24-MAR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/045,957
FILING DATE: 12-APR-1993
ATTORNEY/AGENT INFORMATION:
NAME: Elmer, James Scott
REGISTRATION NUMBER: 36,129
REFERENCE/DOCKET NUMBER: S-19825/PL/CSC 1727
TELECOMMUNICATION INFORMATION:
TELEPHONE: (919)541-8614
TELEFAX: (919)541-8689
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 1201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-456-240-26

Query Match 90.0%; Score 14.4; DB 2; Length 1201;
Best Local Similarity 93.8%; Pred. No. 1.7e+02;
Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 TTGTGCAATTTTC 16
||||| |||||||
Db 152 TTGTGCAATTTTC 137

RESULT 15

US-08-455-736-26/C
 Sequence 26, Application US/08455736
 Patent No. 5880328
 GENERAL INFORMATION:
 APPLICANT: Ryals, John A.
 APPLICANT: Alexander, Danny C.
 APPLICANT: Beck, James J.
 APPLICANT: Duesing, John H.
 APPLICANT: Friedrich, Leslie B.
 APPLICANT: Goodman, Robert M.
 APPLICANT: Harms, Christian
 APPLICANT: Meins, Jr., Frederick
 APPLICANT: Montoya, Alice
 APPLICANT: Moyer, Mary B.
 APPLICANT: Neuhaus, Jean-Marc
 APPLICANT: Payne, George B.
 APPLICANT: Sperison, Christoph
 APPLICANT: Stinson, Jeffrey R.
 APPLICANT: Uknes, Scott J.
 APPLICANT: Ward, Eric R.
 APPLICANT: Williams, Speciosa C.
 TITLE OF INVENTION: CHEMICALLY REGULATABLE AND ANTI-PATHOGENIC
 TITLE OF INVENTION: DNA SEQUENCES AND USES THEREOF
 NUMBER OF SEQUENCES: 106
 CORRESPONDENCE ADDRESS:
 ADDRESS: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: New York
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/455,736
 FILING DATE: 31-MAY-1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/181,271
 FILING DATE: 13-JAN-1994
 APPLICATION NUMBER: US 08/093,301
 FILING DATE: 16-JUL-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/937,197
 FILING DATE: 6-NOV-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/678,378
 FILING DATE: 1-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/305,566
 FILING DATE: 6-FEB-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/165,667
 FILING DATE: 8-MAR-1988
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/042,847
 FILING DATE: 6-APR-1993
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/632,441
 FILING DATE: 21-DEC-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/425,504
 FILING DATE: 20-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/848,506
 FILING DATE: 6-MAR-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/768,122
 FILING DATE: 27-SEP-1991
 PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/580,431
 FILING DATE: 7-SEP-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/368,672
 FILING DATE: 20-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/329,018
 FILING DATE: 24-MAR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/045,957
 FILING DATE: 12-APR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Elmer, James Scott
 REGISTRATION NUMBER: 36,129
 REFERENCE/DOCKET NUMBER: S-19825/P1/CGC 1727
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (919)541-8614
 TELEFAX: (919)541-8689
 INFORMATION FOR SEQ ID NO: 26:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1201 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 US-08-455-736-26

Query Match 90.0%; Score 14.4; DB 2; Length 1201;
 Best Local Similarity 93.8%; Pred. No. 1.7e+02;
 Matches 15; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 TTGTGTCATTTC 16
 ||||| |||||
 Db 152 TTGTGTCATTTC 137

Search completed: January 23, 2003, 20:35:57
 Job time : 29.3429 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using SW model

Run on: January 23, 2003, 19:25:32 ; Search time 136.971 Seconds
(without alignments)
259.276 Million cell updates/sec

Title: US-09-803-736-1022_COPY_10615_10630
Perfect score: 1
Sequence: 1 ttgtgtcaattttc 16

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N.Geneseq.101002:*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	16	100.0	611590	21	AAF22303
2	15	93.8	440	22	ABA12894
3	15	93.8	462	22	AAK67022
4	15	93.8	462	22	AAK67023
5	15	93.8	462	22	AAK67024
6	15	93.8	499	24	AAK95072
7	15	93.8	499	24	AAK95115
8	15	93.8	1609	21	AAA96625
9	15	93.8	1609	21	AAA50828

10	15	93.8	1974	17	AA742918
11	15	93.8	1974	23	AA033338
12	15	93.8	2745	24	ABL28798
13	15	93.8	6254	18	AAV74715
14	15	93.8	15929	22	ABA18496
15	15	93.8	15929	22	ABA19020
16	15	93.8	15964	22	ABA18497
17	15	93.8	15964	22	ABA19021
18	14.4	90.0	241	22	AA117374
19	14.4	90.0	245	22	AA109485
20	14.4	90.0	257	21	AA433124
21	14.4	90.0	375	21	AA229607
22	14.4	90.0	570	22	AA669225
23	14.4	90.0	658	23	AA66758
24	14.4	90.0	711	24	AA69682
25	14.4	90.0	835	21	AA77145
26	14.4	90.0	903	24	AA899448
27	14.4	90.0	907	24	AAV53461
28	14.4	90.0	920	19	AB199279
29	14.4	90.0	1058	22	AA90089
30	14.4	90.0	1078	20	AA20337
31	14.4	90.0	1078	20	AA62822
32	14.4	90.0	1201	20	AAV81615
33	14.4	90.0	1371	23	AA53570
34	14.4	90.0	1438	10	AAV91275
35	14.4	90.0	2247	23	AB126922
36	14.4	90.0	2369	23	AB127146
37	14.4	90.0	2510	23	AB127145
38	14.4	90.0	2616	22	AA53399
39	14.4	90.0	2672	23	AB117494
40	14.4	90.0	2761	23	AB127152
41	14.4	90.0	2990	23	AA43405
42	14.4	90.0	3109	22	AA47627
43	14.4	90.0	3159	22	AB122238
44	14.4	90.0	3198	18	AA792702
45	14.4	90.0			

ALIGNMENTS

RESULT 1
AAF22303
ID AAF22303 standard; DNA: 611590 BP.
XX AAF22303;
XX
XX
XX 20-MAR-2001 (first entry)
XX
XX Arabidopsis thaliana chromosome 2 centromere.
XX
XX Centromere; mitochondrion; vector; ds.
XX
XX Arabidopsis thaliana.
XX
XX WO20005325-A2.
XX
XX 21-SEP-2000.
XX
XX 17-MAR-2000; 2000WO-US07392.
XX
XX 18-MAR-1999; 99US-0125219.
XX PR 01-APR-1999; 99US-0127409.
XX PR 18-MAY-1999; 99US-0134770.
XX PR 13-SEP-1999; 99US-013584.
XX PR 17-SEP-1999; 99US-0154603.
XX
XX (UNCH-) UNIV CHICAGO.
XX
XX Preuss D, Copenhagen G, Keith K;
XX
XX WPI, 2000-587529/55.
XX

SAG13 promoter fro
Arabidopsis thaliana
Drosophila melanog
Staphylococcus aur
Human nervous syst
Human nervous syst
Human nervous syst
Human nervous syst
Human breast cancer
Human secreted exp
Fragmented YAC set
Human cervical can
DNA encoding novel
Streptococcus poly
Human OREF OREF2700
Mouse mRNA for apo
DNA encoding a Sta
Mouse ischemic co
Human bone marrow
Borrelia burgdorferi
Tobacco PR-2 gene
PR-2' protein enc
Helicobacter pylori
DNA encoding Asper
Drosophila melanog
Drosophila melanog
Candida albicans e
Drosophila melanog
Drosophila melanog
Drosophila melanog
Carbolic genomic seq
Candida CarH1 gen
Drosophila melanog
Candida CarH1 gen

PT Recombinant DNA construct comprising a plant centromere, useful for
PT producing stably inherited microchromosomes which can serve as vectors for
PT the construction of transgenic plant and animal cells -
XX
PS Claim 45; Page 820-959; 1449pp; English.
XX
CC The present invention relates to a recombinant DNA construct of a plant
CC (Arabidopsis thaliana) centromere. The constructs are useful for
CC producing stably inherited microchromosomes which can serve as vectors for
CC the construction of transgenic plant and animal cells expressing
CC selected proteins such as hormones, enzymes, interleukins, clotting
CC factors, cytokines, antibodies, and growth factors.
XX
SQ Sequence 611590 BP; 181893 A; 124460 C; 120254 G; 164983 T; 0 other;
Query Match 100.0%; Score 16; DB 21; Length 611590;
Best Local Similarity 100.0%; Pred. No. 2,7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGTCATTTTC 16
Db 228197 TTGTTGTCATTTTC 228212
RESULT 2
ID ABA12894
XX ABA12894 standard; cDNA; 440 BP.
AC ABA12894;
XX
DT 23-JAN-2002 (first entry)
XX
DE Human nervous system related polynucleotide SEQ ID NO 1901.
XX
KW Human; nocotropic; neuroprotective; cytosolic; dermatological; virucide;
KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
KW antiparkinsonian; antisticking; antianemic; antiallergic; cancer;
KW antirheumatic; hepatotropic; cerebroprotective; antifungal;
KW antidiabetic; antidiabetic; anticancer; cardiovascular disorder;
KW antiparasitic; cardiac; immune disorder; neurotropic; gene therapy; vaccine; ss.
XX
OS Homo sapiens.
XX
PN WO200159063-A2.
XX
PD 16-AUG-2001.
XX
PF 17-JAN-2001; 2001MO-DS01334.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0188874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214866.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0218880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 16-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0228287.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 06-SEP-2000; 2000US-0231242.
PR 06-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234272.
PR 25-SEP-2000; 2000US-0234998.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 27-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239937.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.

PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMAN-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI: 2001-541565/60.
 DR P-FSDB; ABR16568.
 PT Nucleic acids encoding 3224 human nervous system antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 XX
 PS Claim 1; SEQ ID NO 1901; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABAI1004-ABR2153) and proteins
 CC (ABAI4678-ABR18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, Rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcl_sequences.
 CC
 XX

SQ Sequence 440 BP; 113 A; 98 C; 103 G; 123 T; 3 other;
 Query Match 93.8%; Score 15; DE 22; Length 440;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTTGTCATTGTT 15
 Db 240 TTGTTGTCATTGTT 254
 RESULT 3
 ID AAK67022 standard; DNA; 462 BP.
 AC AAK67022;
 DT 06-NOV-2001 (first entry)
 DE Human immune/haematopoietic antigen genomic sequence SEQ ID NO:21834.
 DE Human immune/haematopoietic; immune/haematopoietic antigen; cancer;
 KW cytostatic; gene therapy; vaccine; metastasis; ds.
 OS Homo sapiens.
 PN WQ200157182-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01354.
 XX
 PR 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.
 PR 07-JUN-2000; 2000US-0209467.
 PR 28-JUN-2000; 2000US-0214886.
 PR 30-JUN-2000; 2000US-0215135.
 PR 07-JUL-2000; 2000US-0216647.
 PR 07-JUL-2000; 2000US-0216880.
 PR 11-JUL-2000; 2000US-0217487.
 PR 11-JUL-2000; 2000US-0217496.
 PR 14-JUL-2000; 2000US-0218290.
 PR 26-JUL-2000; 2000US-0220963.
 PR 26-JUL-2000; 2000US-0220964.
 PR 14-AUG-2000; 2000US-0224518.
 PR 14-AUG-2000; 2000US-0224519.
 PR 14-AUG-2000; 2000US-0225213.
 PR 14-AUG-2000; 2000US-0225214.
 PR 14-AUG-2000; 2000US-0225216.
 PR 14-AUG-2000; 2000US-0225267.
 PR 14-AUG-2000; 2000US-0225268.
 PR 14-AUG-2000; 2000US-0225270.
 PR 14-AUG-2000; 2000US-0225447.
 PR 14-AUG-2000; 2000US-0225757.
 PR 14-AUG-2000; 2000US-0225758.
 PR 14-AUG-2000; 2000US-0225759.
 PR 18-AUG-2000; 2000US-0226279.
 PR 22-AUG-2000; 2000US-0226681.
 PR 22-AUG-2000; 2000US-0226686.
 PR 22-AUG-2000; 2000US-0227182.
 PR 23-AUG-2000; 2000US-0227009.
 PR 30-AUG-2000; 2000US-0228924.
 PR 01-SEP-2000; 2000US-0229287.
 PR 01-SEP-2000; 2000US-0229343.
 PR 01-SEP-2000; 2000US-0229344.
 PR 05-SEP-2000; 2000US-0229359.
 PR

PR 05-SEP-2000; 2000US-0229513.
 PR 06-SEP-2000; 2000US-0230437.
 PR 06-SEP-2000; 2000US-0230438.
 PR 08-SEP-2000; 2000US-0231242.
 PR 08-SEP-2000; 2000US-0231243.
 PR 08-SEP-2000; 2000US-0231244.
 PR 08-SEP-2000; 2000US-0231413.
 PR 08-SEP-2000; 2000US-0231414.
 PR 08-SEP-2000; 2000US-0231415.
 PR 12-SEP-2000; 2000US-0232081.
 PR 12-SEP-2000; 2000US-0232080.
 PR 14-SEP-2000; 2000US-0232397.
 PR 14-SEP-2000; 2000US-0232398.
 PR 14-SEP-2000; 2000US-0232399.
 PR 14-SEP-2000; 2000US-0232400.
 PR 14-SEP-2000; 2000US-0232401.
 PR 14-SEP-2000; 2000US-0233063.
 PR 14-SEP-2000; 2000US-0233064.
 PR 14-SEP-2000; 2000US-0233065.
 PR 21-SEP-2000; 2000US-0234223.
 PR 21-SEP-2000; 2000US-0234274.
 PR 25-SEP-2000; 2000US-0234497.
 PR 25-SEP-2000; 2000US-0234498.
 PR 26-SEP-2000; 2000US-0235494.
 PR 27-SEP-2000; 2000US-0235834.
 PR 27-SEP-2000; 2000US-0235836.
 PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0237040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 13-OCT-2000; 2000US-0239937.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241221.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-0244617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.

PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249219.
 PR 17-NOV-2000; 2000US-0249224.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249254.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249269.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 06-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PR (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM;
 PI WPI: 2001-483426/52.
 DR Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PS metastasis.
 PS Disclosure; SEQ ID NO 21834; 3071bp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK62170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For
 CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK67694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 462 BP; 125 A; 66 C; 90 G; 181 T; 0 other;
 Query Match 93.8%; Score 15; DB 22; Length 462;
 Best Local Similarity 100.0%; Pred. No. 7.9e-02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTTGTCATTTT 15
 Db 193 TTGTTGTCATTTT 207
 RESULT 4
 AAK67023
 ID AAK67023 standard; DNA: 462 BP.
 XX
 AC AAK67023;
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21835.

XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW Cytostatic; gene therapy; vaccine; metastasis; ds.
XX Homo sapiens.
OS WO200157182-A2.
PN 09-AUG-2001.
PD 17-JAN-2001; 2001WO-US01354.
PF XX
XX 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186550.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225477.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 05-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.

PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX	(HUMA-) HUMAN GENOME SCI INC.	
PA		
PA		
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-483426/52.	
XX		
PT	Nucleic acids encoding human immune/hematopoietic antigen polypeptides,	
PT	useful for preventing, diagnosing and/or treating cancers and	
PT	metastasis	
PS		
XX	Disclosure: SEQ ID NO 21835: 3071pp + Sequence Listing: English.	
XX		
CC	AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)	
CC	amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic	
CC	activity, and can be used in gene therapy and vaccine production. (I)	
CC	proteins and polynucleotides may be used in the prevention, diagnosis and	
CC	treatment of diseases associated with inappropriate (I) expression. For	
CC	example, they may be used to treat disorders associated with decreased	
CC	expression by rectifying mutations or deletions in a patient's genome	
CC	that affect the activity of (I) by expressing inactive proteins or to	
CC	supplement the patients own production of (I). Additionally, (I)	
CC	polynucleotides may be used to produce the secreted (I), by inserting	
CC	the nucleic acids into a host cell and culturing the cell to express the	
CC	protein. (I) proteins and polynucleotides may be used to prevent,	
CC	diagnose and treat immune/hematopoietic-related diseases, especially	
CC	cancers and cancer metastases of haematopoietic-derived cells. AAK64703	
CC	to AAK67694 represent human immune/hematopoietic antigen genomic	
CC	sequences from the present invention. AAK54942 to AAK54950 and AAM82169	
CC	represent sequences used in the exemplification of the present invention.	
XX		
SQ	Sequence 462 BP; 125 A; 66 C; 90 G; 181 T; 0 other;	
	Query Match	93.8%; Score 15; DB 22; Length 462;
	Best Local Similarity	100.0%; Pred. No. 7.9e+02;
	Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0.	
QY	1 TTCTTGCAATTTT 15 	
DB	193 TTGTTGCATTTT 207	
RESULT 5		
AAK67024		
ID	AAK67024 standard; DNA; 462 BP.	
XX		
AC	AAK67024:	
XX		
DT	06-NOV-2001 (first entry)	
XX		
DE	Human immune/hematopoietic antigen genomic sequence SEQ ID NO:21836.	
KX		
KM	Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;	
KW	cytostatic; gene therapy; vaccine; metastasis; ds.	
OS	Homo sapiens.	
XX		
PN	WO200157182-A2.	
PD		
XX	09-AUG-2001.	
PF		
XX	17-JAN-2001; 2001MO-USO1354.	
PR	31-JAN-2000; 2000US-0179065.	
PR	04-FEB-2000; 2000US-0180628.	
PR	24-FEB-2000; 2000US-0184664.	
PR	02-MAR-2000; 2000US-0186350.	
PR	16-MAR-2000; 2000US-0189874.	
PR	17-MAR-2000; 2000US-0190076.	
PR	18-APR-2000; 2000US-0198123.	
PR	19-MAY-2000; 2000US-0205515.	

PR	17-UNU-2000	2000US -0294647
PR	28-JUN-2000	2000US -0214866
PR	30-JUN-2000	2000US -0215135
PR	07-JUL-2000	2000US -0216647
PR	07-JUL-2000	2000US -0216880
PR	11-JUL-2000	2000US -0217446
PR	11-JUL-2000	2000US -0217467
PR	14-JUL-2000	2000US -0218290
PR	14-JUL-2000	2000US -0218963
PR	26-JUL-2000	2000US -0220964
PR	14-AUG-2000	2000US -0224518
PR	14-AUG-2000	2000US -0224519
PR	14-AUG-2000	2000US -0225114
PR	14-AUG-2000	2000US -0225214
PR	14-AUG-2000	2000US -0225267
PR	14-AUG-2000	2000US -0225268
PR	14-AUG-2000	2000US -0225268
PR	14-AUG-2000	2000US -0225270
PR	14-AUG-2000	2000US -0225477
PR	14-AUG-2000	2000US -0225477
PR	14-AUG-2000	2000US -0225757
PR	14-AUG-2000	2000US -0225758
PR	14-AUG-2000	2000US -0225759
PR	18-AUG-2000	2000US -0226279
PR	22-AUG-2000	2000US -0226681
PR	22-AUG-2000	2000US -0226681
PR	22-AUG-2000	2000US -0226682
PR	23-AUG-2000	2000US -0227182
PR	23-AUG-2000	2000US -0227189
PR	30-AUG-2000	2000US -0228924
PR	01-SEP-2000	2000US -0229287
PR	01-SEP-2000	2000US -0229347
PR	01-SEP-2000	2000US -0229345
PR	01-SEP-2000	2000US -0229345
PR	05-SEP-2000	2000US -0229519
PR	05-SEP-2000	2000US -0229519
PR	06-SEP-2000	2000US -0230437
PR	06-SEP-2000	2000US -0230438
PR	08-SEP-2000	2000US -0231242
PR	08-SEP-2000	2000US -0231244
PR	08-SEP-2000	2000US -0231244
PR	08-SEP-2000	2000US -0231413
PR	08-SEP-2000	2000US -0231414
PR	08-SEP-2000	2000US -0232080
PR	08-SEP-2000	2000US -0232080
PR	12-SEP-2000	2000US -0231367
PR	14-SEP-2000	2000US -0232397
PR	14-SEP-2000	2000US -0232398
PR	14-SEP-2000	2000US -0232439
PR	14-SEP-2000	2000US -0232439
PR	14-SEP-2000	2000US -0232440
PR	14-SEP-2000	2000US -0232463
PR	14-SEP-2000	2000US -0233063
PR	14-SEP-2000	2000US -0233065
PR	14-SEP-2000	2000US -0233065
PR	21-SEP-2000	2000US -0234223
PR	21-SEP-2000	2000US -0234223
PR	23-SEP-2000	2000US -0234977
PR	23-SEP-2000	2000US -0234977
PR	25-SEP-2000	2000US -0235498
PR	26-SEP-2000	2000US -0235498
PR	27-SEP-2000	2000US -0235834
PR	27-SEP-2000	2000US -0235834
PR	29-SEP-2000	2000US -0236337
PR	29-SEP-2000	2000US -0236337
PR	29-SEP-2000	2000US -0236366
PR	29-SEP-2000	2000US -0236368
PR	29-SEP-2000	2000US -0236370
PR	02-OCT-2000	2000US -0236602
PR	02-OCT-2000	2000US -0237037
PR	02-OCT-2000	2000US -0237038
PR	02-OCT-2000	2000US -0237040
PR	13-OCT-2000	2000US -0239937
PR	13-OCT-2000	2000US -0239937
PR	20-OCT-2000	2000US -0240960
PR	20-OCT-2000	2000US -0241221

PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0241826.
 PR 01-NOV-2000; 2000US-024617.
 PR 08-NOV-2000; 2000US-0246474.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246610.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249297.
 PR 17-NOV-2000; 2000US-0249299.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250160.
 PR 01-DEC-2000; 2000US-0250391.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251858.
 PR 08-DEC-2000; 2000US-0251859.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-0259678.
 PA (HUMA-) HUMAN GENOME SCI INC.
 PI Rosen CA, Barash SC, Ruben SM;
 DR WPI; 2001-483426/52.
 XX
 PT Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
 PT useful for preventing, diagnosing and/or treating cancers and
 PT metastasis -
 XX
 PS Disclosure: SEQ ID NO 21836; 3071pp + Sequence Listing; English.
 XX
 CC AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)
 CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic
 CC activity, and can be used in gene therapy and vaccine production. (I)
 CC proteins and polynucleotides may be used in the prevention, diagnosis and
 CC treatment of diseases associated with inappropriate (I) expression. For

CC example, they may be used to treat disorders associated with decreased
 CC expression by rectifying mutations or deletions in a patient's genome
 CC that affect the activity of (I) by expressing inactive proteins or to
 CC supplement the patient's own production of (I). Additionally, (I)
 CC polynucleotides may be used to produce the secreted (I), by inserting
 CC the nucleic acids into a host cell and culturing the cell to express the
 CC protein. (I) proteins and polynucleotides may be used to prevent,
 CC diagnose and treat immune/hematopoietic-related diseases, especially
 CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703
 CC to AAK7694 represent human immune/hematopoietic antigen genomic
 CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169
 CC represent sequences used in the exemplification of the present invention.
 XX
 SQ Sequence 462 BP; 125 A; 66 C; 90 G; 181 T; 0 other;
 Query Match 93.8%; Score 15; DB 22; Length 462;
 Best Local Similarity 100.0%; Pred. No. 7.9e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 DB 1 TGTGTGCAATTTT 15
 193 TGTGTGCAATTTT 207
 RESULT 6
 AAK9072
 ID AAK9072 standard; DNA; 499 BP.
 AC AAK9072;
 XX
 DT 24-MAR-2002 (first entry)
 XX
 DE DNA encoding Staphylococcus aureus S21 ribosomal protein gene.
 XX
 KW Staphylococcus aureus ribosomal polypeptide S20; antibacterial;
 KW bacterial ribosomal assembly; food poisoning; multisystem dysfunction;
 KW S21; toxic shock syndrome; skin rash; inhibitor; gene; ds.
 XX
 OS Staphylococcus aureus.
 XX
 PN WO200208265-A2.
 XX
 PD 31-JAN-2002.
 XX
 PE 19-JUL-2001; 2001WO-US21103.
 XX
 PR 19-JUL-2000; 2000US-219361P.
 XX
 PA (PHAA) PHARMACIA & UPJOHN CO.
 XX
 PI Pearson JD, Slightom JL, Chosay JG, McCroskey MC, Shinabarger DJ,
 PI Wilcox S;
 PI Wilecox S;
 DR WPI; 2002-268962/31.
 PT P-PSDB; AAO20067.
 PT Novel isolated Staphylococcus aureus S20 ribosomal polypeptide, useful
 PT for identifying inhibitors of bacterial ribosomal assembly -
 XX
 PS Disclosure; Page 81; 83pp; English.
 XX
 CC The invention relates to an isolated S. aureus ribosomal polypeptide S20,
 CC and the isolated polynucleotide molecules that encode them, vectors and
 CC host cells comprising such polynucleotide molecules and also methods for
 CC the identification of agents that effect ribosomal assembly. The isolated
 CC polypeptide of the invention is useful for identifying inhibitors of
 CC bacterial ribosomal assemblies. The inhibitors identified by the method
 CC of the invention are useful as antibacterial compounds. The antibacterial
 CC compounds can be used against certain strains of S. aureus that can cause
 CC skin rashes, food poisoning, or multisystem dysfunction (toxic shock
 CC syndrome). Fragments of the polynucleotide of the invention are useful as
 CC probes or primers. This polynucleotide sequence represents the DNA
 CC encoding the Staphylococcus aureus S21 ribosomal protein of the

CC Invention.

XX Sequence 499 BP; 202 A; 64 C; 70 G; 163 T; 0 other;

SO Query Match

93.8%; Score 15; DB 24; Length 499;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTGTGCAATTTT 15

DB 395 TTGTGTGCAATTTT 409

RESULT 7

AAK93115

ID AAK93115 standard; DNA; 499 BP.

XX AAK93115;

DT 27-MAY-2002 (first entry)

DE DNA encoding the Staphylococcus aureus S21 ribosomal protein.

XX

KM Staphylococcus aureus ribosomal polypeptide S16; antibacterial;

KW bacterial ribosomal assembly; food poisoning; multisystem dysfunction;

KM S21; toxic shock syndrome; skin rash; inhibitor; gene; ds.

XX

OS Staphylococcus aureus.

XX

PN W0200208266-A2.

XX

PD 31-JAN-2002.

XX

PF 19-JUL-2001; 2001WO-US21104.

XX

PR 19-JUL-2000; 2000US-219360P.

XX

PA (PHMA) PHARMACIA & UPJOHN CO.

XX

PI Pearson JD, Slightom JL, Chosay JG, Shmabarger DL;

XX

DR WPI; 2002-268963/31.

XX

DR P-PSDB; AAO20089.

XX

PT Novel isolated Staphylococcus aureus S16 ribosomal polypeptide, useful

XX

PS Disclosure; Page 72; 73pp; English.

XX

CC The invention relates to an isolated *S. aureus* ribosomal polypeptide S16,

CC

CC and the isolated polynucleotide molecules that encode them, vectors and

CC

CC host cells comprising such polynucleotide molecules and also methods for

CC

CC the identification of agents that effect ribosomal assembly. The isolated

CC

CC polypeptide of the invention is useful for identifying inhibitors of

CC

CC bacterial ribosomal assemblies. The inhibitors identified by the method

CC

CC of the invention are useful as antibacterial compounds. The antibacterial

CC

CC compounds can be used against certain strains of *S. aureus* that can cause

CC

CC skin rashes, food poisoning, or multisystem dysfunction (toxic shock

CC

CC syndrome). Fragments of the polynucleotide of the invention are useful as

CC

CC probes or primers. This polynucleotide sequence represents the DNA

CC

CC encoding the Staphylococcus aureus S21 ribosomal protein of the

XX

SO Sequence 499 BP; 202 A; 64 C; 70 G; 163 T; 0 other;

Query Match 93.8%; Score 15; DB 24; Length 499;

Best Local Similarity 100.0%; Pred. No. 7.9e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTGTGCAATTTT 15

DB 395 TTGTGTGCAATTTT 409

RESULT 8

AAA49625

ID AAA49625 standard; DNA; 1609 BP.

XX AAA49625;

AC AAA49625;

XX 27-OCT-2000 (first entry)

DE Zebrafish parathyroid hormone type-1 receptor PTH1R coding sequence.

XX Zebrafish: parathyroid hormone type-1 receptor; PTH1R;

XX developmental disorder; physiological disorder; neurological disorder;

XX ss.

OS Brachydanio rerio.

XX

FN Key

FT Location/Qualifiers

FT CDS

FT 1..1609

FT /*tag= a

FT /product= "PTH1R"

PN W0200032775-A1.

XX 08-JUN-2000.

XX 30-NOV-1999; 99WO-US28207.

XX 30-NOV-1998; 98US-0110467.

XX (JUEP/) JUEPNER H.

XX (RUBI/) RUBIN D A.

XX Juepner H, Rubin DA;

XX WPI; 2000-412323/35.

XX P-PSDB; AAT9600.

XX

PT New nucleic acids encoding parathyroid hormone receptors PTH1R and

XX PTH3R, useful for treating diseases or disorders associated with

XX impaired receptor functions comprises a specific nucleotide sequence -

XX Claim 2; Fig 2A; 11pp; English.

XX

CC The present sequence is the parathyroid hormone type-1 receptor

CC (PTH1R) coding sequence from the zebrafish. It was obtained by

CC sequencing a cDNA clone. The gene and protein can be used to detect

CC diseases in man where the receptor is either overexpressed or

CC underexpressed, and they can be used to treat these diseases, which may

CC be developmental, physiological or neurological disorders. They can also

CC be used to identify agonists and antagonists which can be used in a

CC similar manner. In addition, the gene can be used for chromosome

Query Match 93.8%; Score 15; DB 21; Length 1609;

Best Local Similarity 100.0%; Pred. No. 8e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTGTGCAATTTT 15

DB 974 TTGTGTGCAATTTT 988

RESULT 9

AAA30828

ID AAA30828 standard; DNA; 1609 BP.

XX AAA30828;

XX 29-AUG-2000 (first entry)

DE Zebrafish PTH1R receptor coding sequence.
 XX
 XX Zebrafish; PTH1R receptor; PTH3R receptor; diagnosis; cancer;
 KW parathyroid hormone type 1 receptor; parathyroid hormone type 3 receptor;
 KW ds.
 XX
 OS Brachydanio rerio.
 XX
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1608
 FT /*tag= a
 FT /product= PTH1R receptor
 FT /partial
 FT /note= "no stop codon given"
 FT
 XX
 PN MO20032771-A1.
 XX
 XX 08-JUN-2000.
 PD
 XX
 PF 28-MAY-1999; 99WO-US11883.
 XX
 PF 30-NOV-1998; 98US-0110467.
 XX
 XX (JUEP/) JUEPNER H.
 PA (RUBI/) RUBIN D A.
 XX
 P1 Juepner H, Rubin DA;
 XX
 DR WPI: 2000-412319/35.
 DR P-PEDB; AA90230.
 XX
 XX Novel zebrafish parathyroid hormone/parathyroid hormone related peptide
 PT receptor 3 and isolated nucleic acid encoding zebrafish parathyroid
 PT hormone receptor 1 for treating disorders associated with receptor
 PT function
 XX
 PS Claim 2; Fig 1c; 11pp; English.
 XX
 CC This sequence encodes a parathyroid hormone receptor type 1 (PTH1R)
 CC receptor protein of the invention. The invention also relates to a PTH3R
 CC receptor protein. Antagonists of PTH1R or PTH3R can be used for the
 CC treatment of diseases associated with an increase in PTH3R or PTH1R
 CC activity, respectively. The peptides are used for diagnosis or prognosis
 CC of diseases and disorders associated with PTH1R or PTH3R, such as cancer.
 CC The polypeptides can be used as a molecular weight markers on sodium
 CC dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels, or
 CC on molecular sieve gel filtration columns. Antigenic epitope-bearing
 CC peptides and polypeptides are useful to raise antibodies, including
 CC monoclonal antibodies, that bind specifically to a polypeptide. The
 CC peptides are useful during diagnosis of diseases and disorders in
 CC mammals involving PTH1R or PTH3R receptor expression or function.
 CC Mutations that affect PTH1R or PTH3R sequence and/or expression levels
 CC of PTH1R or PTH3R could be diagnostic for patients with disease or
 CC disorders of a developmental, physiological or neurological nature. The
 CC nucleic acid molecules are valuable for chromosome identification. The
 CC mapping of DNAs to chromosomes is an important first step in
 CC correlating those sequences with genes associated with disease.
 XX
 SQ Sequence 1609 BP; 409 A; 400 C; 397 G; 403 T; 0 other;
 Query Match 93.8%; Score 15; DB 21; Length 1609;
 Best Local Similarity 100.0%; Pred. No. 8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTTGTCATTTT 15
 ||||||||||||
 DB 974 TTGTTGTCATTTT 988
 RESULT 10
 AAT42918
 ID AAT42918 standard; DNA: 1974 BP.
 XX

AC AAT42918;
 XX
 XX 18-JAN-1997 (first entry)
 XX
 DE SAG13 promoter from senescence-associated gene.
 XX
 XX SAG13; promoter; senescence-associated gene; development;
 KW Arabidopsis thaliana; gene regulation; 5'-untranslated region;
 KW cytokinin biosynthesis; isopentenyltransferase; transgenic plant;
 KW flowering; seed; fruit; crop improvement; ds.
 XX
 OS Arabidopsis thaliana.
 XX
 XX
 FH Key Location/Qualifiers
 FT Promoter 1..1782
 FT /*tag= a
 FT variation 1009
 FT /*tag= b
 FT 5'UTR 1782..1956
 FT /*tag= c
 XX
 PN MO9629858-A1.
 XX
 XX 03-OCT-1996.
 PD
 XX
 PF 20-FEB-1996; 96WO-US02313.
 XX
 PF 28-MAR-1995; 95US-0413135.
 XX
 XX (WISC) WISCONSIN ALUMNI RES FOUND.
 FA
 XX Amasino RM, Gan S;
 XX WPI: 1996-454877/45.
 XX
 XX Senescence associated gene promoters, SAG12 and SAG13, - useful for
 PT producing genetic constructs for producing transgenic plants having
 PT delayed senescence
 XX
 PS Claim 3; Page 29-30; 38pp; English.
 XX
 CC The sequence represents an SAG13 promoter from an Arabidopsis
 CC thaliana senescence-associated gene. The sequence also contains
 CC the SAG13 gene 5'-untranslated region. The promoter is present in
 CC 2 copies in the A. thaliana genome, with an A-G variation at
 CC position 1009. The sequence has all regulatory signals required for
 CC senescence-specific gene expression, and may be operably linked
 CC to a cytokinin biosynthesis isopentenyltransferase gene in a vector
 CC for expression in a transgenic plant. The resulting transgenic
 CC plant shows delayed senescence, and shows longer vegetative growth,
 CC producing more flowers, seeds or fruit.
 XX
 SQ Sequence 1974 BP; 654 A; 287 C; 329 G; 703 T; 1 other;
 Query Match 93.8%; Score 15; DB 17; Length 1974;
 Best Local Similarity 100.0%; Pred. No. 8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTTGTCATTTT 15
 ||||||||||||
 DB 1115 TTGTTGTCATTTT 1129
 RESULT 11
 AAD33338
 ID AAD33338 standard; DNA: 1974 BP.
 XX
 AC AAD33338;
 XX
 XX 01-JUN-2002 (first entry)
 XX
 DE Arabidopsis thaliana SAG13 promoter.
 XX

KM Senescence associated gene; SAG-13; transgenic plant; transgenic;
 XX Senescence-specific promoter; senescence characteristic; ds.
 XX Arabidopsis thaliana.
 OS
 XX US6359197-B1.
 PN
 XX
 PD 19-MAR-2002.
 XX
 PF 17-NOV-1997; 97US-0971395.
 XX
 PR 29-MAR-1995; 95US-0413135.
 XX
 PA (MISC) WISCONSIN ALUMNI RES FOUND.
 XX
 PI Amasino RM, Gan S, Noh Y;
 XX
 DR WPI; 2002-291012/33.
 XX
 PT Novel senescence associated promoter sequence connected to a
 XX protein-coding DNA sequence useful for the creation of transgenic
 PT plants with altered senescence characteristics -
 XX
 PS Example; Column 19-22; 21pp; English.
 XX
 CC The invention relates to a genetic construct comprising senescence
 CC specific promoter, preferably senescence associated gene (SAG)-12
 CC promoter operably connected to a protein-coding DNA sequence not
 CC naturally connected to the promoter. The senescence associated
 CC promoters are useful for the creation of transgenic plants with
 CC altered senescence characteristics. Genetic constructs can be
 CC inserted into plants which become effective only upon plant cells
 CC entering senescence. For example, a gene encoding a biosynthetic
 CC enzyme under the control of a senescence-specific promoter can be
 CC inserted into a plant, without having the tissues of the plant
 CC exposed to the excess of cytokinin during pre-senescence growth.
 CC Then at the onset of senescence, the senescence-specific promoter
 CC activates cytokinin production to alter the progression of senescence
 CC in the plant. The present sequence is Arabidopsis thaliana senescence
 CC specific promoter, SAG13.
 XX
 SQ Sequence 1974 BP; 654 A; 287 C; 329 G; 703 T; 1 other;
 XX
 Query Match 93.8%; Score 15; DB 24; Length 1974;
 Best Local Similarity 100.0%; Pred. No. 8e+02; Indels 0; Gaps 0;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTGTCATTTT 15
 DB 1115 TTGTGTCATTTT 1129
 XX
 RESULT 12
 ABL28798/C
 ID ABL28798 standard; DNA; 2745 BP.
 XX
 AC ABL28798;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 37867.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KM Pharmacological; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX

PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-0614150.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX
 PS Claim 1; SEQ ID NO 37867; 21pp + Sequence Listing; English.
 XX
 CC The invention relates to an isolated nucleic acid detection reagent
 CC capable of detecting 1000 or more genes from Drosophila. The invention is
 CC useful in developmental biology and in elucidating cell signalling and
 CC cell-cell interactions in higher eukaryotes for the development of
 CC insecticides, therapeutics and pharmaceutical drugs. The invention
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
 CC sequences (ABR57737-ABR72072).
 CC
 CC The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.
 XX
 SQ Sequence 2745 BP; 800 A; 618 C; 593 G; 734 T; 0 other;
 XX
 Query Match 93.8%; Score 15; DB 23; Length 2745;
 Best Local Similarity 100.0%; Pred. No. 8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTGTCATTTT 15
 DB 829 TTGTGTCATTTT 815
 XX
 RESULT 13
 AAV74715
 ID AAV74715 standard; DNA; 6254 BP.
 XX
 AC AAV74715;
 XX
 DT 16-MAR-1999 (first entry)
 XX
 DE Staphylococcus aureus contig SEQ ID #404.
 XX
 KW Computer readable medium; vaccine; S. aureus infection; immunodetection;
 KM cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KM skin infection; surgical wound infection; scalded skin syndrome;
 KM toxic shock syndrome; ds.
 XX
 OS Staphylococcus aureus.
 XX
 FH Key Location/Qualifiers
 FT misc_feature 1561..1620
 FT /tag= a
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT misc_feature 3361..3420
 FT /tag= b
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"
 FT misc_feature 5161..5220
 FT /tag= c
 FT /note= "these bases represent a line of missing text in
 FT the sequence listing in the specification. They
 FT are included to maintain the nucleotide numbering
 FT given in the specification for this DNA sequence"

FT XX given in the specification for this DNA sequence"
PN XX EF786519-A2.
XX PD 30-JUL-1997.
XX PF 07-JAN-1997; 97EP-0100117.
XX PR 05-JAN-1996; 96US-0009861.
XX PA (HUMA-) HUMAN GENOME SCI INC.
XX PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
XX PI Rosen CA;
XX DR WPI; 1997-374922/35.
XX PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
XX PT stored on computer readable medium and used in the production of
XX PT anti-*S.aureus* vaccines
XX PS Claim 1; Page 1297-1300; 3271pp; English.
XX CC This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences
XX CC of the invention. The DNA sequences are recorded on a computer readable
XX CC medium, preferably selected from a floppy or hard disk, random access
XX CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using
XX CC the *S.aureus* DNA sequences allows putative functions to be assigned so
XX CC that protein-encoding or regulatory regions of commercial, therapeutic or
XX CC industrial importance can be obtained. Specifically, sequences which are
XX CC likely to encode antigens have been identified and these polypeptides can
XX CC be used in a vaccine composition against *S.aureus* infection. The
XX CC polypeptides can also be used in a kit for the immunodetection of
XX CC *S.aureus* in a sample. *S.aureus* is implicated in numerous human diseases,
XX CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,
XX CC skin and surgical wound infections, scaled skin syndrome, toxic shock
XX CC syndrome, etc. Organisms transformed with the DNA sequences can be used
XX CC for recombinant production of the polypeptides. The new DNA sequences
XX CC (and their fragments) are useful as primers or probes for isolating
XX CC homologues of any of the *S.aureus* DNA sequences contained on the
XX CC computer readable medium.
SQ Sequence 6254 BP; 2215 A; 822 C; 1171 G; 1852 T; 194 other;
Query Match 93.8%; Score 15; DB 18; Length 6254;
Best local similarity 100.0%; Pred No. 8e+02;
Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGTTGTCATTTTT 15
DB 216 TTGTTGTCATTTTT 230
|||||
RESULT 14
ID ABA18496 standard; DNA; 15929 BP.
XX ABA18496;
XX 23-JAN-2002 (first entry)
XX DE Human nervous system related polynucleotide SEQ ID NO 10827.
XX XX Human; nootropic; neuroprotective; cyrostatic; dermatological; vituicide;
XX KW immunosuppressive; anti-inflammatory; anti-HIV; antibacterial; vulnerary;
XX KW antiparkinsonian; antiskinking; antianemic; antiarthritic; cancer;
XX KW antihemetic; hepatotropic; cerebroprotective; antiinflammatory;
XX KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;
XX KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
XX OS Homo sapiens.
XX XX

PN WO200159063-A2.
XX 16-AUG-2001.
XX PF 17-JAN-2001; 2001WO-US01334.
XX PR 31-JAN-2000; 2000US-0179065.
XX PR 04-FEB-2000; 2000US-0180628.
XX PR 24-FEB-2000; 2000US-0184664.
XX PR 02-MAR-2000; 2000US-0186350.
XX PR 16-MAR-2000; 2000US-0189874.
XX PR 17-MAR-2000; 2000US-0190076.
XX PR 18-APR-2000; 2000US-0198123.
XX PR 19-MAY-2000; 2000US-0205515.
XX PR 07-JUN-2000; 2000US-0209467.
XX PR 28-JUN-2000; 2000US-0214886.
XX PR 30-JUN-2000; 2000US-0215135.
XX PR 07-JUL-2000; 2000US-0216647.
XX PR 07-JUL-2000; 2000US-0216880.
XX PR 11-JUL-2000; 2000US-0217487.
XX PR 11-JUL-2000; 2000US-0217496.
XX PR 14-JUL-2000; 2000US-0218290.
XX PR 26-JUL-2000; 2000US-0220963.
XX PR 26-JUL-2000; 2000US-0220964.
XX PR 14-AUG-2000; 2000US-0224518.
XX PR 14-AUG-2000; 2000US-0224519.
XX PR 14-AUG-2000; 2000US-0225213.
XX PR 14-AUG-2000; 2000US-0225214.
XX PR 14-AUG-2000; 2000US-0225266.
XX PR 14-AUG-2000; 2000US-0225267.
XX PR 14-AUG-2000; 2000US-0225268.
XX PR 14-AUG-2000; 2000US-0225270.
XX PR 14-AUG-2000; 2000US-0225447.
XX PR 14-AUG-2000; 2000US-0225757.
XX PR 14-AUG-2000; 2000US-0225758.
XX PR 14-AUG-2000; 2000US-0225759.
XX PR 18-AUG-2000; 2000US-0226279.
XX PR 22-AUG-2000; 2000US-0226681.
XX PR 22-AUG-2000; 2000US-0226868.
XX PR 22-AUG-2000; 2000US-0227182.
XX PR 23-AUG-2000; 2000US-0227182.
XX PR 30-AUG-2000; 2000US-0228924.
XX PR 01-SEP-2000; 2000US-0229287.
XX PR 01-SEP-2000; 2000US-0229343.
XX PR 01-SEP-2000; 2000US-0229344.
XX PR 01-SEP-2000; 2000US-0229345.
XX PR 05-SEP-2000; 2000US-0229509.
XX PR 05-SEP-2000; 2000US-0229513.
XX PR 06-SEP-2000; 2000US-0230437.
XX PR 06-SEP-2000; 2000US-0230438.
XX PR 08-SEP-2000; 2000US-0231242.
XX PR 08-SEP-2000; 2000US-0231243.
XX PR 08-SEP-2000; 2000US-0231244.
XX PR 08-SEP-2000; 2000US-0231413.
XX PR 08-SEP-2000; 2000US-0231414.
XX PR 08-SEP-2000; 2000US-0232080.
XX PR 08-SEP-2000; 2000US-0232081.
XX PR 12-SEP-2000; 2000US-0231968.
XX PR 14-SEP-2000; 2000US-0232397.
XX PR 14-SEP-2000; 2000US-0232398.
XX PR 14-SEP-2000; 2000US-0232399.
XX PR 14-SEP-2000; 2000US-0232400.
XX PR 14-SEP-2000; 2000US-0232401.
XX PR 14-SEP-2000; 2000US-0233063.
XX PR 14-SEP-2000; 2000US-0233064.
XX PR 14-SEP-2000; 2000US-0233065.
XX PR 21-SEP-2000; 2000US-0234223.
XX PR 21-SEP-2000; 2000US-0234274.
XX PR 25-SEP-2000; 2000US-0234997.
XX PR 25-SEP-2000; 2000US-0234998.
XX PR 26-SEP-2000; 2000US-0235484.
XX PR 27-SEP-2000; 2000US-0235834.
XX PR 27-SEP-2000; 2000US-0235836.

PR 29-SEP-2000; 2000US-0236327.
 PR 29-SEP-2000; 2000US-0236367.
 PR 29-SEP-2000; 2000US-0236368.
 PR 29-SEP-2000; 2000US-0236369.
 PR 29-SEP-2000; 2000US-0236370.
 PR 02-OCT-2000; 2000US-0236802.
 PR 02-OCT-2000; 2000US-0237037.
 PR 02-OCT-2000; 2000US-0237038.
 PR 02-OCT-2000; 2000US-0237039.
 PR 13-OCT-2000; 2000US-0239040.
 PR 13-OCT-2000; 2000US-0239935.
 PR 20-OCT-2000; 2000US-0240960.
 PR 20-OCT-2000; 2000US-0241785.
 PR 20-OCT-2000; 2000US-0241786.
 PR 20-OCT-2000; 2000US-0241787.
 PR 20-OCT-2000; 2000US-0241808.
 PR 20-OCT-2000; 2000US-0241809.
 PR 20-OCT-2000; 2000US-0242221.
 PR 01-NOV-2000; 2000US-0246417.
 PR 08-NOV-2000; 2000US-0246475.
 PR 08-NOV-2000; 2000US-0246476.
 PR 08-NOV-2000; 2000US-0246477.
 PR 08-NOV-2000; 2000US-0246478.
 PR 08-NOV-2000; 2000US-0246523.
 PR 08-NOV-2000; 2000US-0246524.
 PR 08-NOV-2000; 2000US-0246525.
 PR 08-NOV-2000; 2000US-0246526.
 PR 08-NOV-2000; 2000US-0246527.
 PR 08-NOV-2000; 2000US-0246528.
 PR 08-NOV-2000; 2000US-0246532.
 PR 08-NOV-2000; 2000US-0246609.
 PR 08-NOV-2000; 2000US-0246611.
 PR 08-NOV-2000; 2000US-0246613.
 PR 17-NOV-2000; 2000US-0249207.
 PR 17-NOV-2000; 2000US-0249208.
 PR 17-NOV-2000; 2000US-0249209.
 PR 17-NOV-2000; 2000US-0249210.
 PR 17-NOV-2000; 2000US-0249211.
 PR 17-NOV-2000; 2000US-0249212.
 PR 17-NOV-2000; 2000US-0249213.
 PR 17-NOV-2000; 2000US-0249214.
 PR 17-NOV-2000; 2000US-0249215.
 PR 17-NOV-2000; 2000US-0249216.
 PR 17-NOV-2000; 2000US-0249217.
 PR 17-NOV-2000; 2000US-0249218.
 PR 17-NOV-2000; 2000US-0249244.
 PR 17-NOV-2000; 2000US-0249245.
 PR 17-NOV-2000; 2000US-0249264.
 PR 17-NOV-2000; 2000US-0249265.
 PR 17-NOV-2000; 2000US-0249267.
 PR 17-NOV-2000; 2000US-0249269.
 PR 17-NOV-2000; 2000US-0249300.
 PR 01-DEC-2000; 2000US-0250391.
 PR 01-DEC-2000; 2000US-0251160.
 PR 05-DEC-2000; 2000US-0251030.
 PR 05-DEC-2000; 2000US-0251988.
 PR 05-DEC-2000; 2000US-0256719.
 PR 06-DEC-2000; 2000US-0251479.
 PR 08-DEC-2000; 2000US-0251856.
 PR 08-DEC-2000; 2000US-0251868.
 PR 08-DEC-2000; 2000US-0251869.
 PR 08-DEC-2000; 2000US-0251989.
 PR 08-DEC-2000; 2000US-0251990.
 PR 11-DEC-2000; 2000US-0254097.
 PR 05-JAN-2001; 2001US-02595678.
 XX (HUMA-) HUMAN GENOME SCI INC.
 PA Rosen CA, Barash SC, Ruben SM.
 XX
 PI

XX WPI; 2001-541565/60.
 DR Nucleic acids encoding 324 human nervous system antigen polypeptides,
 XX useful for preventing, diagnosing and/or treating nervous system
 PT cancers and metastases -
 PS Disclosure: SEQ ID NO 10827; 1701pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABA11004-ABA21534) and proteins
 CC (ABA14678-ABA18001) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pcr_sequences.
 XX
 SQ Sequence 15929 BP; 4592 A; 3191 C; 3342 G; 4804 T; 0 other;
 Query Match 93.8%; Score 15; DB 22; Length 15929;
 Best Local Similarity 100.0%; Pred. No. 8e+02;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTGTCATTTT 15
 Db 7209 TTGTGTCATTTT 7223
 RESULT 15
 ABA19020
 ID ABA19020 standard; DNA; 15929 BP.
 AC ABA19020;
 XX
 XX 23-JAN-2002 (first entry)
 DE Human nervous system related polynucleotide SEQ ID NO 11351.
 XX
 KW Human; nootropic; neuroprotective; cytostatic; dermatological; virocidic;
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;
 KW antiparkinsonian; antiskinking; antianaemic; antiarthritic; cancer;
 KW antithaumatic; hepatotropic; cerebroprotective; antiinflammatory;
 KW antiallergic; antidiabetic; antitumor; anticonvulsant; antitungal;
 KW antiparasitic; cardiac; immune disorder; cardiovascular disorder;
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.
 OS Homo sapiens.
 XX
 XX WO200159063-A2.
 XX
 PD 16-AUG-2001.
 XX
 PF 17-JAN-2001; 2001WO-US01334.
 XX
 XX 31-JAN-2000; 2000US-0179065.
 PR 04-FEB-2000; 2000US-0180628.
 PR 24-FEB-2000; 2000US-0184664.
 PR 02-MAR-2000; 2000US-0186350.
 PR 16-MAR-2000; 2000US-0189874.
 PR 17-MAR-2000; 2000US-0190076.
 PR 18-APR-2000; 2000US-0198123.
 PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.
PR 23-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225271.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239335.
PR 13-OCT-2000; 2000US-0239337.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241785.

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 20-OCT-2000; 2000US-0242221.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250391.
PR 01-DEC-2000; 2000US-0251160.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251858.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.

PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM:
XX WPI, 2001-541565/60.
XX
DR Nucleic acids encoding 3224 human nervous system antipeptides,
XX useful for preventing, diagnosing and/or treating nervous system
PT cancers and metastases -
PT
XX
PS Disclosure: SEQ ID NO 11351; 1701pp + Sequence Listing: English.
XX
CC The invention relates to novel genes (AB11004-AB21534) and proteins
CC (AB14678-AB18001) useful for preventing, treating or ameliorating
CC medical conditions e.g. by protein or gene therapy. The genes are
CC isolated from a range of human tissues disclosed in the specification.
CC The nucleic acids, proteins, antibodies and (ant)agonists are useful

CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX

Sequence 15929 BP; 4592 A; 3191 C; 3342 G; 4804 T; 0 other;

Query Match 93.8%; Score 15; DB 22; Length 15929;

Best Local Similarity 100.0%; Pred. No. 8e+02;

Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

0Y 1 TTGTTGTCATTTT 15

DB 7209 TTGTTGTCATTTT 7223

Search completed: January 23, 2003, 21:27:10
 Job time : 197.971 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 23, 2003, 19:24:08 ; Search time 1226.06 Seconds
(without alignments)
379.791 Million cell updates/sec

Title: US-09-803-736-1022_COPY_10615_10630

Perfect score: 16

Sequence: 1 ttgtgtcaatttttc 16

Scoring table: IDENTITY_NDC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : GenBank.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sv.*
13: gb_un.*
14: gb_vt.*
15: gb_yt.*
16: em_ba.*
17: em_fun.*
18: em_hum.*
19: em_in.*
20: em_mu.*
21: em_om.*
22: em_ov.*
23: em_pal.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_vt.*
30: em_yt.*
31: em_hgt_hum.*
32: em_hgt_inv.*
33: em_hgt_other.*
34: em_hgt_mus.*
35: em_hgt_pln.*
36: em_hgt_rtd.*
37: em_hgt_man.*
38: em_hgt_vtl.*
39: em_sy.*
40: em_hgt_hum.*
41: em_hgt_mus.*
42: em_hgt_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	16	100.0	1850	3	AB072595	AB072595 Halocynthia
C 2	16	100.0	2310	6	AX488817	AX488817 Sequence
C 3	16	100.0	11396	1	AE010523	AE010523 Fusobacte
C 4	16	100.0	11851	1	AE006341	AE006341 Lactococcc
C 5	16	100.0	27364	3	U41026	U41026 Caenorhabd
C 6	16	100.0	39841	9	AC113353	AC113353 Homo sapi
C 7	16	100.0	42355	3	CEP26F2	CEP26F2
C 8	16	100.0	48128	6	AX059515	AX059515 Sequence
C 9	16	100.0	93711	9	AC091880	AC091880 Homo sapi
C 10	16	100.0	94302	8	AC007730	AC007730 Arabidops
C 11	16	100.0	101445	2	AC093973	AC093973 Rattus no
C 12	16	100.0	142396	9	AL157407	AL157407 Human DNA
C 13	16	100.0	149834	2	AC105683	AC105683 Rattus no
C 14	16	100.0	161658	2	AC117244	AC117244 Mus muscu
C 15	16	100.0	164405	9	AC016256	AC016256 Homo sapi
C 16	100.0	168267	8	AP003535	AP003535 Oryza sat	
C 17	16	100.0	171125	9	AC023457	AC023457 Homo sapi
C 18	16	100.0	180334	9	AC129471	AC129471 Homo sapi
C 19	16	100.0	189672	9	CNS01DYM	AL136059 Human chr
C 20	16	100.0	191923	9	AC087886	AC087886 Homo sapi
C 21	16	100.0	196459	2	AC010986	AC010986 Homo sapi
C 22	16	100.0	210949	2	AC127610	AC127610 Rattus no
C 23	16	100.0	213353	2	AC010258	AC010258 Homo sapi
C 24	16	100.0	244698	2	AC097054	AC097054 Rattus no
C 25	16	100.0	270852	2	AC125346	AC125346 Mus muscu
C 26	16	100.0	301662	2	AC127291	AC127291 Mus muscu
C 27	15	93.8	499	6	AX369409	AX369409 Sequence
C 28	15	93.8	499	6	AX380860	AX380860 Sequence
C 29	15	93.8	546	5	AF069394	AF069394 Gallus ga
C 30	15	93.8	647	8	EC002836	U02636 Eleusine co
C 31	15	93.8	1855	5	AF033669	AF033669 Gallus ga
C 32	15	93.8	1962	8	CACDC10G	Z25870 C. albicans
C 33	15	93.8	1974	6	AR201518	AR201518 Sequence
C 34	15	93.8	1974	6	I75098	I75098 Sequence
C 35	15	93.8	2129	5	AF132084	AF132084 Dario rer
C 36	15	93.8	3260	3	AY122131	AY122131 Drosophi
C 37	15	93.8	3344	3	DMDCT112	X16866 Drosophila
C 38	15	93.8	3344	3	DMDRC112	X16864 Drosophila
C 39	15	93.8	3344	3	DMDRC1V2	X16968 Drosophila
C 40	15	93.8	3344	3	DMDRC1C2	X16971 Drosophila
C 41	15	93.8	3532	8	AF192276	AF192276 Arabidops
C 42	15	93.8	6377	3	AB050668	AB050668 Branchios
C 43	15	93.8	17591	9	AC018916	AC018916 Homo sapi
C 44	15	93.8	19483	3	CEP19A6	CEP19A6
C 45	15	93.8	26168	3	CEP08G12	Z66571 Caenorhabd

ALIGNMENTS

RESULT 1	AB072595/c	1850 bp	mRNA	linear	INV 06-APR-2002
LOCUS	AB072595	1850 bp	mRNA	linear	INV 06-APR-2002
DEFINITION	Halocynthia roretzi HtWnt-5beta mRNA for ascidian homolog of wnt-5, complete cds.				
ACCESSION	AB072595				
VERSION	AB072595.1	GI:20065724			
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
1					

AUTHORS Miya, T. and Nishida, H.
 TITLE Isolation of cDNA clones for mRNAs transcribed zygotically during cleavage stage in ascidian, *Halocynthia roretzi*
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1850)
 AUTHORS Miya, T. and Hiroki, N.
 TITLE Direct Submission
 JOURNAL Submitted (05-OCT-2001) Takahito Miya, Tokyo Institute of Technology, Department of Biological Sciences, 4259, Nagatsuta, Midori-ku, Yokohama, Kanagawa 226-8501, Japan
 (E-mail: tmiya@bio.titech.ac.jp, Tel: 81-45-924-5721, Fax: 81-45-924-5722)

FEATURES

source
 Location/Qualifiers
 1..1850
 /organism="Halocynthia roretzi"
 /db_xref="taxon:7729"
 /clone_lib="cDNA library of 110-cell embryo mRNAs
 subtracted with fertilized egg mRNAs"
 /dev_stage="110-cell embryo"
 1..1850
 /gene="HrMnt-5beta"
 136..1251
 /gene="HrMnt-5beta"
 /codon_start=1
 /product="ascidian homolog of wnt-5"
 /protein_id="BAB88819.1"
 /db_xref="GI:20065725"
 /translation="MPQSAVFLISCHVLEFAVITKTNQWFGSTANARATVA
 DTSIPSEISLSHSGRAICOKYRDHLLVSEAGGIECOYFRRARNNCTTNK
 KRKIDNDSSTVGVGVIDIGSRAPFYAISAGVYHAIARGCKRGRLACGCSNPR
 PDGIRDEMGGCGGNLDYAYAGFAHETIDAFEDNSPNNKSSRRVYMLHNHNGR
 MTVRGARPTCKGNSGSCINTCWLRVHREIDKIRKHSSEILVYNNRSMR
 PRKKSISPTIDLVHDSPPDYCRANSRGVLTGRCPCNHSDDSGGLMCCGRG
 YDTRVDESPRCRCFQWCVRCQCKIKRQRYROR"

BASE COUNT

557 a 375 c 422 g 496 t

ORIGIN

Query Match 100.0%; Score 16; DB 3; Length 1850;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 TTGTTGCAATTTTTC 16
 ||||||||||||
 DB 461 TTGTTGCAATTTTTC 446

RESULT 2

AX488817 2310 bp DNA linear PAT 16-AUG-2002
 LOCUS AX488817
 DEFINITION Sequence 6117 from Patent WO2053728.
 ACCESSION AX488817
 VERSION AX488817.1 GI:22322829

KEYWORDS

Candida albicans.
 Candida albicans.
 Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 Saccharomycetales; mitosporeic Saccharomycetales; Candida.

REFERENCE

1 Roemer, T., Jiang, B., Boone, C., Bussey, H. and Ohlsen, K. L.
 TITLE Gene disruption methodologies for drug target discovery
 JOURNAL Patent: WO 02053728-A 6117 11-JUL-2002;
 Elitra Pharmaceuticals, Inc. (US)

FEATURES

source
 Location/Qualifiers
 1..2310
 /organism="Candida albicans"
 /db_xref="taxon:5476"
 BASE COUNT 943 a 316 c 270 g 781 t
 ORIGIN

Query Match 100.0%; Score 16; DB 6; Length 2310;
 Best Local Similarity 100.0%; Pred. No. 1.4e+03;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTTGCAATTTTTC 16
 ||||||||||||
 DB 2047 TTGTTGCAATTTTTC 2032

RESULT 3

AE010523 11396 bp DNA linear BCF 25-MAR-2002
 LOCUS AE010523
 DEFINITION Fusobacterium nucleatum subsp. nucleatum ATCC 25586, section 65 of 197 of the complete genome.
 ACCESSION AE010523 AE009951
 VERSION AE010523.1 GI:19713508
 KEYWORDS

SOURCE

Fusobacterium nucleatum subsp. nucleatum ATCC 25586.
 Fusobacterium nucleatum subsp. nucleatum ATCC 25586
 Bacteria; Fusobacteria; Fusobacterium.

REFERENCE

1 (bases 1 to 11396)
 Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T.,
 Lykidis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G.,
 Zhu, L., Vasileva, O., Chu, L., Kogan, Y., Chaga, O., Goldsman, E.,
 Bernal, A., Larsen, N., D'Souza, M., Walunas, T., Pusch, G.,
 Haselkorn, R., Fouts, M., Kyriades, N. and Overbeek, R.
 Genome sequence and analysis of the oral bacterium *Fusobacterium*
nucleatum strain ATCC 25586
 J. Bacteriol. 184 (7), 2005-2018 (2002)

TITLE

J. Bacteriol. 184 (7), 2005-2018 (2002)

JOURNAL

MEDLINE

PUBMED

REFERENCES

2 (bases 1 to 11396)
 Kapral, V., Anderson, I., Ivanova, N., Reznik, G., Los, T.,
 Lykidis, A., Bhattacharyya, A., Bartman, A., Gardner, W., Grechkin, G.,
 Zhu, L., Chu, L., Kogan, Y., Chaga, O., Goldsman, E., Bernal, A.,
 Larsen, N., D'Souza, M., Walunas, T., Pusch, G.D., Haselkorn, R.,
 Fouts, M., Kyriades, N. and Overbeek, R.
 Direct Submission
 Submitted (13-FEB-2002) Integrated Genomics, 2201 W. Campbell Park
 Drive, Chicago, IL 60612, USA
 Location/Qualifiers

FEATURES

source
 Location/Qualifiers
 1..11396
 /organism="Fusobacterium nucleatum subsp. nucleatum ATCC
 25586"
 /strain="ATCC 25586"
 /sub_species="nucleatum"
 /db_xref="ATCC:25586"
 /db_xref="taxon:190304"
 84..737
 /gene="FN0081"
 84..737
 /gene="FN0081"
 /codon_start=1
 /transl_table=11
 /product="Ethanolamine utilization protein eutL"
 /protein_id="AA194294.1"
 /db_xref="GI:19713509"
 /translation="MNDPLKRSVLSVKLIIPNDKMAEELNIPNGRSIGIITASD
 DVTYALDENATKAAEVIVYAKSEYGGANANAKLAGVEIGIMAGPNAEYKSGLNA
 IDFIENGCGYSSANEDDIIPYACVSRGTSGISKTAGYEDDAALAYIAIPPEAMIA
 IDALKKADVYTAAPFQPSFETNRGGILTGSSACKSACDAFAAVAFVQNPXKI"
 748..1194
 /gene="FN0082"
 748..1194
 /gene="FN0082"
 /codon_start=1
 /transl_table=11
 /product="Ethanolamine utilization protein eutM"
 /protein_id="AA194295.1"
 /db_xref="GI:19713510"
 /translation="MKALGLIETRGWGAIVAADILKTAQVLEINERKVGGLWIE
 FEGDVAVKASVAAVATLKDGIYGVSHVIPRDSVSKIIKRNASEEVEVIERKV
 KEIKKTEIOEESTETLKNIEEIEINELIKYSKKTKTKK"
 1229..1513
 /gene="FN0083"
 1229..1513
 /gene="FN0083"

```

gene
CDS
/codon_start=1
/transl_table=11
/product="ethanolamine utilization protein eutM precursor"
/protein_id="AAL94296.1"
/db_xref="GI:19713511"
/transl_table="MSTLALGMIEFTKGLVAAYEADAMKANTYLGKEIYGGGLV
TWMVGVGAVKATIDAGAAADRVGELISVHVLRPHSEVELLPKSN"
1621..3063
/gene="FN0084"
1621..3063
/gene="FN0084"
/EC_number="1.2.1.10"
/codon_start=1
/transl_table=11
/product="Acetaldehyde dehydrogenase [acetylating]"
/protein_id="AAL94297.1"
/db_xref="GI:19713512"
/transl_table="MDDRLISIQVREDLIRKAKVAQKLYSTFTQEDIDKRVCAIYEEK
KHHYDLAKKANEETFGKCEDVIRKKNANEVYDCIKDKMTVGILNETDVTISVG
PMGIYAALPSTNPTSTAIYKTLISLGNALIVSPHPNAKNCVIDTVKIMQKAAVAA
GAPVGLIGVITPIETGNETLMSKSDTSILATGGEAMVRAAYSSGRPAIGSPNGP
AYTEKANTIKKAVKTIETSTPNNGVYCSQSVVEPCKEAVMDEFPKGGFPLSK
ESEKLGRIIRPMTNPOLIVKDKQTLAKIAGLNIPIANKYILSEONTVSKTPYS
REKLTITLAFYVEENAKACERAIEELLEWGBGHTLIHSDDKILIRERALKPKYSRM
LINVGSIGVGATNLAPAPLTGCGAVGSSSTSDITPMNLINRIVAGVRELSDP
KKDITNSLDVINSEVEDMIRRIAEYR"
3319..4086
/gene="FN0085"
3319..4086
/gene="FN0085"
/EC_number="2.5.1.17"
/codon_start=1
/transl_table=11
/product="Ethanolamine utilization cobalamn
adenosyltransferase"
/protein_id="AAL94298.1"
/db_xref="GI:19713513"
/transl_table="MWLSEDLIKYRKPPVPEIEKGLTLTPSAKQFLNEKGIEL
IKEEKPLVSTKNEEDVNAEKKIPEKPKYVKNDEYFEKPEYMWVADNLIIFENS
KLIALRGKIEPLAVLTGKEIETLSNDKLIIDETIIFKIONIVAAKELKLIEN
CTEPDSKIDKIKETIEMPEKPKGHLLIETLSNDLTIHKLNRRLRLARELEIQAD
FYEDYKVSRKDLFAENILSDVIYVILKVDNGEYR"
4086..4673
/gene="FN0086"
4086..4673
/gene="FN0086"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"
/protein_id="AAL94299.1"
/db_xref="GI:19713514"
/transl_table="MNNPDENYITELVKKELSRVLTDOGIEIKREYVFLDDHEIKEQ
LSQKPNFSNAKTLIVSOLSKNLNYSNAYENYEKRIKFLLEKREYIIRKEIE
YSKVENIPLAVOKKEEYLEKIKSGIKYKNDPYNISLMQKEEYVKKLIDNKIKE
LEAKGMRITIVENSITITSSABETAKENIETIKR"
4673..4923
/gene="FN0087"
4673..4923
/gene="FN0087"
/codon_start=1
/transl_table=11
/product="Ethanolamine utilization protein eutN"
/protein_id="AAL94300.1"
/db_xref="GI:19713515"
/transl_table="MLIGEVIGWAKRKYDGLGKFLIYKTEDNKRMYAFDSVAG
IGEKVILISGSSARNAIMRDVYVDAAIIGIIDGMOEE"
4936..5271
/gene="FN0088"
4936..5271
/gene="FN0088"
/codon_start=1
/transl_table=11
/product="Hypothetical protein"

```

```

gene
CDS
/protein_id="AAL94301.1"
/db_xref="GI:19713516"
/transl_table="MSKRPTITKNEEDIGSGKITYIDKRAIISSSLDYIREHNIQV
YEEFCSAKPSEDCALKEEVNNTNSRADPAGAVAMVVKILINDYIGIQEKKIMQV
IKILIEVLK"
5271..6353
/gene="FN0089"
5271..6353
/gene="FN0089"
/codon_start=1
/transl_table=11
/product="Ethanolamine permease"
/protein_id="AAL94302.1"
/db_xref="GI:19713517"
/transl_table="MGINETIITYMVFMANVGAADKCTGKRFYGGKFEETMAGAL
ALSNMIVASLAPVANTLRPVGPVYVAAAGADPAPEATITLLANDKGYPLASLADP
MYGKFPAGILISMGAGATVFTIPVALGIIIEKEDRPTAKGVAAIFPGCEVGLV
AGPPVAILRNILPILIFAVLITIGLMPPEKMTGTEFGTCVVVITDGLAAITR
NLTGVILIPGMAPIGEGMITSIAIVLAGAPLVHFTVFKKPLEKYEGLGNEI
GAAGVYASLANNIPFGMMDMPNCKVNVANVACAFVGGHLPFTGGVYDKAMTAP
MNGKLAGGILALITAKILTTTKAR"
6364..6813
/gene="FN0090"
6364..6813
/gene="FN0090"
/codon_start=1
/transl_table=11
/product="Ethanolamine utilization protein eutQ"
/protein_id="AAL94303.1"
/db_xref="GI:19713518"
/transl_table="MDKELSEELIRKVIKEELGAQESSEHYKQMDKSGIVGVKLNOM
KKRVKMDNPNDOVYTTTDLFTTQESRPGAGIMKMTETTPMVLTVYEDLVYIIGRL
EIIIDRKVYGEAGVITLIERKSKIEFSPVNAKRLIYVYANMSL"
6837..7937
/gene="FN0091"
6837..7937
/gene="FN0091"
/EC_number="3.1.3.3"
/codon_start=1
/transl_table=11
/product="Phosphoserine phosphatase"
/protein_id="AAL94304.1"
/db_xref="GI:19713519"
/transl_table="MSTENSCVRIDEGRNPNKRRVLEKILKRYDYNVAAVEDMNT
STGDTQGLNLFITYQENTIKYKNDKQENYIKRNPYDDEKRYKNDGLBLNVTKA
NDIYKDIPLIENITSDKSLKKEIRTEPRFPAKHHYIYHNLDPGSAELACIME
FYLLSGMTKDEKVSIAKAVDTFKGELIGVAVESRYVTEBAGVETIYDNGLRIRP
EMANLYHELRKNGIDVYIISASQELIEVAFNDSSYVNLVENVYAKLKSTTDNITL
IDRYNDIPIPTQRECKSETINKFKSYNGRGLVAGDAVGDESMILTKPEDTEVILLI
MKREGKLDVAKDSRALIQKNAGTGLDPRKN"
7957..9075
/gene="FN0092"
7957..9075
/gene="FN0092"
/EC_number="1.1.1.-"
/codon_start=1
/transl_table=11
/product="NADPH-dependent butanol dehydrogenase"
/protein_id="AAL94305.1"
/db_xref="GI:19713520"
/transl_table="MKFPRIQPKILFGEEDSLDYLTKLCKKKVMIVTDEVMQLRLTDF

```

Query Match 100.0%; Score 16; DB 1; Length 11396;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTCAATTTTTC 16
|||||
Db 2641 TTGTGTCAATTTTTC 2626

RESULT 4
AE006341

LOCUS	AEO06341	11851 bp	DNA	linear	BCF 14-MAY-2001
DEFINITION	Lactococcus lactis subsp. lactis ILL403 section 103 of 218 of the complete genome.				
ACCESSION	AEO06341	AEO05176			
VERSION	AEO06341.1	GI:12724039			
KEYWORDS	Lactococcus lactis subsp. lactis.				
SOURCE	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;				
ORGANISM	Lactococcus.				
REFERENCE	1 (bases 1 to 11851)				
AUTHORS	Bojotin,A., Wincker,P., Manger,S., Jaillon,O., Malarme,K., Weissenbach,J., Ehrlich,S.D. and Sorokin,A.				
TITLE	The complete genome sequence of the lactic acid bacterium Lactococcus lactis ssp. lactis ILL403				
JOURNAL	Genome Res. 11 (5), 731-753 (2001)				
MEDLINE	21235186				
PUBMED	11337471				
REFERENCE	2 (bases 1 to 11851)				
AUTHORS	Bojotin,A., Wincker,P., Manger,S., Jaillon,O., Malarme,K., Weissenbach,J., Ehrlich,S.D. and Sorokin,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (09-JAN-2001) INRA, Genetique Microbiennne, Domaine de Vilvert, Jouy en Josas 78352, France				
FEATURES	Location/Qualifiers				
source	1..11851				
	/organism="Lactococcus lactis subsp. lactis"				
	/strain="ILL403"				
	/sub_species="lactis"				
	/db_xref="taxon:1360"				
gene	82..1050				
	/gene="phol"				
	/note="L92886"				
CDS	82..1050				
	/gene="phol"				
	/note="EVIDENCE BY HOMOLOG BIO14.01 Adaptations and atypical conditions"				
	/codon_start=1				
	/transl_table=11				
	/product="phosphate starvation inducible protein"				
	/protein_id="AAK05182.1"				
	/db_xref="GI:12724040"				
	/translation="MEIYVLRNLEKLNPEPDAISLGVDQYIKIVSGOIGVEIN LYGEFVOISEDETERAKALALVOLVHGCVPHASDVITAIKAVIANDDIENNTA LYEVELTKDAGKSIRLNKLQKTGYSDHVIYFGPATGTEPLAVVAOVOKR NGQVKRIILTRPAVENAGSLPTPGDOEKVPDYLPAYDALFOILKTAETEMERG IIEIAPLAYRGRTLDADAFVLIDEAQNTTMOKMFLFTLGFNSKMTVNDISIDLP NKVSSGLIDKKLSHVSKSIDPVYFAKADVRRPVAAFIKAYGDE"				
gene	1108..1587				
	/gene="ykjE"				
	/note="L93856"				
CDS	1108..1587				
	/gene="ykjE"				
	/note="EVIDENCE BY HOMOLOG BIO15.02 HYPOTHETICAL. Conserved, there are two paralogs in L.lactis, also yggC. Probably ADP-ribose pyrophosphatase or mutT, evidenced by cognitor."				
	/codon_start=1				
	/transl_table=11				
	/product="conserved hypothetical protein"				
	/protein_id="AAK05183.1"				
	/db_xref="GI:12724041"				
	/translation="MAEGYVMDLRKTIKGHPVYIAQASLIIDENRGIILOKADNG KCHYGSEPNKEVEBAAREIXECYKAGKINITVVASGTGEOHEFPYNDDEVHYD IVFTNDSSGIIIESEVLDCCMFAPNLPEDILIATGPIILSCMKMLOKADRK" 1739..2227				
gene	1739..2227				
	/gene="ykjF"				
	/note="L94534"				
CDS	1739..2227				
	/gene="ykjF"				
	/note="EVIDENCE BY HOMOLOG BIO15.01 HYPOTHETICAL.General"				
	/codon_start=1				
	/transl_table=11				
gene	2211..2666				
	/gene="gkx"				
	/note="L95012"				
CDS	2211..2666				
	/gene="gkx"				
	/codon_start=1				
	/transl_table=11				
	/product="diacylglycerol kinase (EC 2.7.1.107)"				
	/protein_id="AAK05185.1"				
	/db_xref="GI:12724043"				
	/translation="MOSGDYKILFKKDNENGRKPVEDERRARKNPILSSDFALS GIMTFKERMRKHALLAEIATVFCIYPOISREPMIFLISITLIPNAELFNALIN VVDIASDYFYMRKRKAAGAVLVISGALLVGLTVLRWHLFF"				
gene	2680..3258				
	/gene="ykjH"				
	/note="L95481"				
CDS	2680..3258				
	/gene="ykjH"				
	/note="EVIDENCE BY HOMOLOG BIO15.01 HYPOTHETICAL.General"				
	/codon_start=1				
	/transl_table=11				
	/product="HYPOTHETICAL PROTEIN"				
	/protein_id="AAK05186.1"				
	/db_xref="GI:12724044"				
	/translation="MDRKNDPLKILKNSELMTILDENVNLNPNYLAAGSVPOTWA NYSDGNLMITGHIDIVTFDKETSADSSAKDKRLKYSEFPYOPDIRHEAMHL WHGNKRPYKNTENAERIALVHAIVGAGISGNSLTFAPGLIEDIFTIKTPITHVD NNRILYENKVARMQDERFSNLKIEMSDELKML"				
gene	complement(3292..3342)				
	/gene="comPC"				
	/note="L0319"				
CDS	complement(3292..3342)				
	/gene="comPC"				
	/note="EVIDENCE BY HOMOLOG BIO04.06 Transformation. ATP-dependent RNA/DNA helicase involved in DNA transport."				
	/codon_start=1				
	/transl_table=11				
	/product="competence protein ComPC"				
	/protein_id="AAK05187.1"				
	/db_xref="GI:12724045"				
	/translation="MHICLLCSNINQNFUSELFLKSPQNLCSCCKNREKIPFE HCPCCKPWHICIDCONMWERKGYRIRHOALFHNQAMMDYFSYKRVGDYRLYQIE EYPFNISQKSPVIVIPISPRKLEERGNOVAFLODNFTELLEKNVSKSSINRRR ERLESPNPRLRKLGKIVPTVILLIDITYTTGTTVHASOILKEAGVHEIRSPSLCR" complement(3939..5261)				
gene	complement(3939..5261)				
	/gene="comPA"				
	/note="L0318"				
CDS	complement(3939..5261)				
	/gene="comPA"				
	/note="EVIDENCE BY HOMOLOG BIO04.06 Transformation. ATP-dependent RNA/DNA helicase involved in DNA transport."				
	/codon_start=1				
	/transl_table=11				
	/product="competence protein ComPA"				
	/protein_id="AAK05188.1"				
	/db_xref="GI:12724046"				
	/translation="KSTNOERLEFRLILKNDILQILKNTDKISVSKIFSNLETFEPRK PLAGMTSISNKKIKNRCGTGVIHAKNSVKLPIDVYCPGQVIOAGRGVADIVYELI ELPKYSINNSGKLIENKOSISNALCOEINSHOVAVPCDAGKTEMIVOEIOL E GVAAGLASPRIDVPCLELHORISPRDFSKIPDLTHGDSYFPAPLIINTSHOILPRKEE FDLIIIDEDVADPFRONEMILVAEKARKIEGNLIYDAISDILKEDILKOKLIPUL LPRFHNPELVVYKPFKNKDKDKLIEQRNSGFPILIFPAALIEGQEFKAOQLKLPLEK IELASVASTKDLRLIVAKFNRKEITILATISILRGVTPENVDVINSEHNPNTKTKS ALIOMAGRVGRSPERPDTGLVSEFHYGSKAMQAVREIKKNMOLGCSF" 5318..5947				
gene	5318..5947				
	/gene="ykjI"				


```

CDS
/note="L98095"
5318..5947
/gene="yk1j1"
/note="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.General"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05189.1"
/db_xref="GI:12724047"
/translation="MTITIKDFINEEKRSFCHIKRISSEPRARDTAKIKKH
WKANWCSAYTLGDSQETQRTSDGEPSTAGVPELTKKEINCAVYTRRGCT
KIGAGGLIAYAGSYNAIEAVGIVFQRLINLDYNNLSLQRFIDRIKISD
SEPLSDIKKCKPIDEKVEVMNLTETPNKRIIMKKDSOOVEERLS"
6081..6578
/gene="yk1j1"
/note="L9876"
6081..6578
/gene="yk1j1"
/note="EVIDENCE BY HOMOLOG Y BIO15.01 HYPOTHETICAL.
General. 41% identical to telococcus resistance protein
of transposon Tn1546 of Enterococcus faecium"
/codon_start=1
/transl_table=11
/product="HYPOTHETICAL PROTEIN"
/protein_id="AAK05190.1"
/db_xref="GI:12724048"
/translation="MTKRLAQLFFTYIIGLVVLYEKLSPHPAPLEMTNRSINLT
PLMAGSGREVFNILAFIPGILLMNAKPSLNKLAISILSFESLQYLLAIG
ATDITPLINSGALGIGISFYLLIKVSKAVDILITPTILLPTIIFRQSVIL
GVAVV"
6626..7645
/gene="yk1j1"
/note="L9942"
6626..7645
/gene="yk1j1"
/note="EVIDENCE BY HOMOLOG Y BIO15.02 HYPOTHETICAL.
Conserved. integral membrane protein"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK05191.1"
/db_xref="GI:12724049"
/translation="MKIPRGSFVSLALITAFVYAGNGSLFITLLAVLEISLNDN
AIVARILRMSFAMOKAFILIGVLAIVGMILPILIVGSAODIPSAIKLAEK
GMPVYGYCYILHHRPQIAEGMFLMLAGFPDPRHMTKIPKTIQKIGN
PANAITSITILLTSTFADRAVAFRGIGILFVLDVGEETSSKANTSS
VVAIGAGALFLYLEVIDASFSDVIGAFATDPTIILIGLVGAGAFRSLLY
LVOKGILNDIVYLEHGAHMAITLAFILASIKWEIGEVYGLIGAVFITLFSIVI
YNRNH"
7755..8507
/gene="nuca"
/note="L100541"
7755..8507
/gene="nuca"

Query Match 100.0%; Score 16; DB 1; Length 11851;
Best Local Similarity 100.0%; Pred. No. 1.1e+03;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTGTCAATTTTC 16
|||||
Db 2460 TTGTGTCAATTTTC 2475

RESULT 5
LOCUS U41026 27364 bp DNA linear INV 12-JUN-2002
DEFINITION Caenorhabditis elegans cosmid C28G1, complete sequence.
ACCESSION U41026
VERSION U41026.1 GI:1086701
KEYWORDS HTG.
SOURCE Caenorhabditis elegans.
ORGANISM Caenorhabditis elegans.

```

REFERENCE

1 (bases 1 to 27364)
 Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida;
 Rhabditidae; Rhabditidae; Feloderinae; Caenorhabditis.
 Waterston, R.
 Genome sequence of the nematode *C. elegans*: a platform for
 investigating biology. The *C. elegans* Sequencing Consortium
 Science 282 (5396), 2012-2018 (1998)
 9851916
 2 (bases 1 to 27364)
 Favello, A.
 The sequence of *C. elegans* cosmid C28G1
 Unpublished (2001)
 3 (bases 1 to 27364)
 Waterston, R.
 Direct Submission
 Submitted (20-NOV-1995) Robert Waterston
 4 (bases 1 to 27364)
 Waterston, R.
 Direct Submission
 Submitted (14-SEP-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 5 (bases 1 to 27364)
 Waterston, R.
 Direct Submission
 Submitted (05-OCT-2001) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 6 (bases 1 to 27364)
 Waterston, R.
 Direct Submission
 Submitted (18-MAY-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 7 (bases 1 to 27364)
 Waterston, R.
 Direct Submission
 Submitted (12-JUN-2002) Department of Genetics, Washington
 University, Genome Sequencing Center, 4444 Forest Park Avenue, St.
 Louis, MO 63110, USA
 Submitted by:
 Genome Sequencing Center
 Department of Genetics, Washington University
 St. Louis, MO 63110, USA, and
 Sanger Centre, Hinxton Hall
 Cambridge CB10 1HQ, England
 email: rwenematode.wustl.edu and jee@sanger.ac.uk

REFERENCE

COMMENT

NOTICE: This sequence may not be the entire insert of this clone.
 It may be shorter because we only sequence overlapping sections
 once, or longer because we provide a small overlap between
 neighboring submissions.

This sequence was finished as follows unless otherwise noted: all
 regions were double stranded, sequenced with an alternate chemistry
 or covered by high quality data (i.e., phred quality >= 30); an
 attempt was made to resolve all sequencing problems, such as
 compressions and repeats; all regions were covered by sequence from
 more than one m13 subclone.

For a graphical representation of this cosmid sequence and its
 analysis see:
<http://www.wormbase.org/db/seq/sequence?name=C28G1;class=Sequence>

NEIGHBORING COSMID INFORMATION

The 5' cosmid is R04E5, 200 bp overlap; the 3' cosmid is C06E2,
 3200 bp overlap. Actual start of this cosmid is at base position 1
 of C28G1; actual end is at 16099 of C06E2.

NOTES:

Coding sequences below are the result of integration and manual review of the following data: computer analysis using the program GeneFinder (P. Green and L. Hillier, personal communication), the large scale EST projects of Yoji Kohara (http://www.ddbj.nig.ac.jp/c-elegans/html/CE_INDEX.html) and The C. elegans ORFome cloning project (<http://world.dfc.harvard.edu/>), similarity to other proteins from Blastx analyses (<http://blast.wustl.edu/>), sequence conservation with C. briggsae using Jim Kent's WABA alignment program (Genome Research 10:1115-1125, 2000), individual C. elegans GenBank submissions, and personal communications with C. elegans researchers. tRNAs are predicted using the program tRNAscan-SE (Lowe, T.M. and Eddy, S.R., 1997, Nucl. Acids. Res., 25, 955-964).

FEATURES

SOURCE

1. 27364

/organism="Caenorhabditis elegans"

/strain="Bristol N2"

/db_xref="taxon:6239"

/chromosome="X"

/clone="C28G1.5"

/complement(421..3167)

/gene="C28G1.5"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=C28G1.5;class=Sequence>"

complement(join(421..482,956..1229,1284..1329,1375..1562,1805..1989,3014..3167))

/gene="C28G1.5"

/note="coded for by the following C. elegans cDNAs:

cm10e3"

/codon_start=1

/product="Hypothetical protein C28G1.5"

/protein_id="AA051522.1"

/db_xref="GI:21392654"

/translation="MNVSAHYHOIYLNTRINGSLVTCLEFSEIKSVPRDYCIYKMI LKRLPYGCHALCHKCTITAMVNNSTVPCPCPTVITVNNDTITLKKKLFALIEYED ARSLNKADVTYSCSCKMCRCEFGNKHKTGLLQEPAAAMRTVLNSETKL ENFCDSIQCSEIYKNAORFESSGATYKAKIEHFKKIKENYACREKTTVLKDA DCCIDNKKKEIISVGHOFKOKLIDMKEFHNINQAGLLSKESQIMELVEMAKIATQ SNVYNAPOPTKMDITINLASFNGK"

complement(3621..4819)

/gene="C28G1.5"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=C28G1.6;class=Sequence>"

complement(join(3621..3679,3729..4002,4052..4106,4157..4344,4416..4481,4528..4703,4750..4819))

/gene="C28G1.6"

/note="coded for by the following C. elegans cDNAs:

cm10e3, yk1342c05.5, yk1038g12.3, yk1342c05.3"

/codon_start=1

/product="Hypothetical protein C28G1.6"

/protein_id="AA051523.1"

/db_xref="GI:21392655"

/translation="MILPCEICEESYRESRIPKINCGHVICPYCASKLPENDEYK CPICPQPLEVDRDVNNLVNFMGLKITIEIESSKKKIIDIPAGPWCAHAEVAKFI CVNTIDYKCLAKRCEFEFGDKRKLRLDEYNMKALIKKSGRSKRTIDLEN SLDEFEKENSLSDSNAFMTINELTNGYDSMRKNDVREKSIIEGLIIEERLME NRRKVGALRSQHYREDKLAEIKRLSLDESLINENGINKSMADTESLVONPSANT PLELSHNVKVELPKPOFC"

complement(5485..9337)

/gene="C28G1.3"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=C28G1.3;class=Sequence>"

complement(join(5485..5593,5921..6037,6086..6694,6741..6880,7058..7211,7259..7380,7427..7753,7831..7924,8103..8173,8215..8256,8307..8613,8701..9031,9265..9337))

/gene="C28G1.3"

/note="contains similarity to Pfam domain PF00097 (zf-C3HC4), Score=29.7, E-value=2.7e-08, N=2; coded for by the following C. elegans cDNAs: yk161d07.5, yk545c12.5, yk1003a11.3, yk761d07.3, yk545c12.3, yk1003a11.3,

yk1169e06.3, yk1185c05.3, yk1185c05.5, yk1169e06.5,

yk1003a11.5"

/codon_start=1

/product="Hypothetical protein C28G1.3"

/protein_id="AA02447.2"

/db_xref="GI:21392653"

/translation="MPNSSISOPATGSETASTSVYVYVMSAEQEPFLYELETDS GNSGIVLITPYDMGVQSFALQORISHDKRIKOVCSFYQSVFAMQELMKLEQ CODIKREYVIAIAEIOQISQALCKKREIYRKRIMKNAKNAQIVACVLENKAK LQEQEMKRYIYALKTLELHSHLALVEKRYTQYLAKSNAPRIETIKKAYSEKRN FLENIKIVAGRIGIKHASKDRRTIFONASRKLTAQSHSGVTAETAEKALQSEAKN ASNEIKIVSADGSIYKNNSPKRNKNOVDDEQSAQDLIDPVPNHCQILRNVLGA KVEEYVYRQOREODIYIPTKNNMFHYVEYEDIEVGFVEQDILMTQSNIST TYDDKRLMDALMKIRHNDARFGSPDYIMMKKVIILFETLMSGYAAYPIYE ILQFRQYNEIYKRYCAQEPRELDKONTPTVYSEEPETILRKFPYKRMQOE PEPREPFSPVIVDATTOKNTILGCKLPNDNIQUTSVDDTVYRANVLRMGV LKSPVHRSLMDIOLVITNLGLESKESLGAFTSKTSGEALGITSROVYLSKVY PDRVSEVEQOIDECKMRKSVDEIIDLANYPWEI.PAAAGASERISDILNFOYTFISF TNLPSGLAKHVCYQTCRHISQSMSPFLISPEYICISTGALDQFSLDVQCMEFTYRCP VAGVDPTLSMFPADLRDLDMNSDMFTTFNENAYEKDHAKEYLRYKASTAIYVLEKMI EPERKSTGPRGIAKGRKILDTYRQLKLET"

complement(9340..9341)

/note="SL1 trans-splice acceptor; see yk1003a11.5"

9589..11286

/gene="C28G1.2"

/note="for a graphical representation of this gene see: <http://www.wormbase.org/db/seq/sequence?name=C28G1.2;class=Sequence>"

complement(join(9589..9754,10330..10547,10592..10738,10828..10914,10960..11112,11155..11286))

/gene="C28G1.2"

/codon_start=1

/product="Hypothetical protein C28G1.2"

/protein_id="AA082351.1"

/db_xref="GI:1086703"

/translation="MPLHCRMAPSRGCHSTFLCAIRGSPILSPSIALVTAQVLLK LKMYVSPWMSVGCQSYVSPISWLPFALGNEKNSIVDPKYSIDGLDSVRSY LKNDLTLSYRADSQINDVRNSPNOTILFLNAGMEKILAEKRSMLFPSPKNSVIT YLIEKIVILNVDDAFOMTSIYKESKISQFVYLVPSQGLKALKLTLLKQFENKLEQ ESTVOLQVIMPKLDIOLCILNISKYLGQPIKISIKYKESRISKMLYRENLPFH FPAIDPVEFVALRNHGPVTFGVFS"

9738..9808

/gene="C28G1.2"

/product="tRNA-Glu"

complement(11963..12034)

/product="tRNA-Gly"

18069..21649

misc.feature

gene

CDS

trna

trna

gene

Query Match

Best Local Similarity 100.0%; Score 16; DB 3; Length 27364;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY

1

TTGTTGCAATTTTC

16

DB

21778

TTGTTGCAATTTTC

21763

RESULT 6

AC113353/c

LOCUS

AC113353

DEFINITION

Homo sapiens chromosome 5 clone CTD-2638F23, complete sequence.

ACCESSION

AC113353

VERSION

AC113353.2

KEYWORDS

HTG

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

1 (bases 1 to 39841)

REFERENCE

Direct Submission

Unpublished

2 (bases 1 to 39841)

TITLE

JOURNAL

REFERENCE

AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (01-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 39841)
 AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
 TITLE Direct Submission
 JOURNAL Submitted (30-MAR-2002) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Mar 30, 2002 this sequence version replaced gi:19033467.
 Draft Sequence Produced by DOE Joint Genome Institute
 www.jgi.doe.gov
 Finishing Completed at Stanford Human Genome Center
 www.shgc.stanford.edu
 Quality: Phrap Quality >=40 100% of Sequence;
 Estimated Total Number of Errors is 0.

FEATURES
 source
 1..39841
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /chromosome="5"
 /clone="CPD-2638P23"

BASE COUNT 10990 a 8134 c 8501 g 12216 t
 ORIGIN

Query Match 100.0%; Score 16; DB 9; Length 39841;
 Best Local Similarity 100.0%; Pred. No. 8.8e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTCAATTTTTC 16
 Db 20310 TTGTGTCAATTTTTC 20295

RESULT 7
 CEP26F2 42355 bp DNA linear INV 23-JAN-2002
 LOCUS Caenorhabditis elegans cosmid F26F2, complete sequence.
 DEFINITION 281514
 ACCESSION 281514.1 GI:3869232
 VERSION HTG.
 KEYWORDS Caenorhabditis elegans.
 SOURCE Caenorhabditis elegans.
 ORGANISM Rhabdicolidae, Rhabdilitidae; Pelodetrinae; Caenorhabditis.
 REFERENCE 1
 AUTHORS none.
 TITLE genome sequence of the nematode C. elegans: a platform for investigating biology. The C. elegans Sequencing Consortium
 JOURNAL Science 282 (5396), 2012-2018 (1998)
 MEDLINE 99069613
 PUBMED 9851916
 REMARK The C. elegans Sequencing Consortium.
 2 (bases 1 to 42355)
 Ainscough, R.
 DIRECT SUBMISSION
 Submitted (06-NOV-1996) Nematode Sequencing Project, Sanger Institute, Hinxton, Cambridge CB10 1SA, England and Department of Genetics, Washington University, St. Louis, MO 63110, USA. E-mail: jess.sanger.ac.uk or rtw@nematoe.wustl.edu
 On Nov 13, 1998 this sequence version replaced gi:2814485.
 Coding sequences below are predicted from computer analysis, using predictions from GeneFINDER (P. Green, U. Washington), and other available information.
 Current sequence finishing criteria for the C. elegans genome sequencing consortium are that all bases are either sequenced unambiguously on both strands, or on a single strand with both a dye primer and dye terminator reaction, from distinct subclones. Exceptions are indicated by an explicit note.
 This sequence is the entire insert of clone F26F2. The true right end of clone Y38H6 is at 11519 in this sequence. The start of this sequence (1..105) overlaps with the end of sequence AU031630.
 The end of this sequence (36344..42355) overlaps with the start of

sequence 281540.
 [981006 dll] : Cosmid flipped
 For a graphical representation of this sequence and its analysis see: - <http://wormbase.sanger.ac.uk/perl/ace/elegans/seq/sequence?name=F26F2>
 IMPORTANT: This sequence is NOT necessarily the entire insert of the specified clone. It may be shorter because we only sequence overlapping sections once, or longer because we arrange for a small overlap between neighbouring submissions.
 Location/Qualifiers
 1..42355
 /organism="Caenorhabditis elegans"
 /db_xref="taxon:6239"
 /chromosome="V"
 /clone="F26F2"
 gene
 join(complement(580..750), complement(297..536), complement(AU031630.1:63775..63852), complement(AU031630.1:63540..63731), complement(AU031630.1:63389..63490), complement(AU031630.1:63254..63344), complement(AU031630.1:63121..63206))
 /gene="F26F2.1"
 join(complement(580..750), complement(297..536), complement(AU031630.1:63775..63852), complement(AU031630.1:63540..63731), complement(AU031630.1:63389..63490), complement(AU031630.1:63254..63344), complement(AU031630.1:63121..63206))
 /gene="F26F2.1"
 /codon_start=1
 /protein_id="CAB82214.1"
 /db_xref="GI:7321099"
 /db_xref="SPRMBL:09N6L0"
 /translation="METGYTPEADQRRKRNKACQENQNDPMTNGDFEQRNSFDSQGERLRGDERHLEONRASRNDQNRKREKREQNPAPVQRRGASSTYNGEYVPRRLTHERPGRGHENRRTSLERPKYNNYNNRQGVFDQYQRAENRQMSDQQWYTWEGQRRKHOTRVGNRNGPTRGNSHTSSADFLHQRQTNNSTQINQOQDHFVNDKRTKSYNGPDKDNPRLKLSQDEINIDGFMQRKMRNNRMLLEADHAKRPAKSIKKEKIKKHTSPTFRGTFTNFFLLVHLQIP"
 join(2955..3029,3123..3222,4653..4750,4798..4959)
 /gene="F26F2.2"
 join(2955..3029,3123..3222,4653..4750,4798..4959)
 /gene="F26F2.2"
 /note="Predicted using GeneFINDER"
 /codon_start=1
 /protein_id="CAB04188.1"
 /db_xref="GI:3876425"
 /db_xref="SPRMBL:09XV57"
 /translation="MSDNTLPAEMLAKONSROHDPMKRMENYNARLSQADRRMARENNLKYVNAEDSKPTNMSDNLTBRABELDKQNSQNDPKATYEMNNASALISQARDNRNATRSNNIKETYNKLVEKQKSPVDISIKTR"
 complement(join(5703..5813,5858..5917,6109..6369,6415..6834))
 /gene="F26F2.3"
 complement(join(5703..5813,5858..5917,6109..6369,6415..6834))
 /gene="F26F2.3"
 /note="Predicted using GeneFINDER"
 /codon_start=1
 /protein_id="CAB04189.1"
 /db_xref="GI:3876426"
 /db_xref="SPRMBL:09YV56"
 /translation="MEQNDISGCTPRITLWQGRGSVSHYQDEGTDIQVVRGSAKIDGVPPHKLITPTDPTRDGLKLTAKNGGGLIKRNVTSRSDQTSRSHATDQRTSFTIPYPTPLGVLKIPYSAAKKNVQIKMRKSSHYRRAQIVGLNPNQHIARLSEESORLQDEARLAEORRNRNDRNDOSTRVVQPREDDNSNSGCHERNQQRREKCIIMREPRYNGCDQTERPMRNPALAEQAASDAQRRMBREELNFIQRF"
 join(10009..10083,10128..10289)
 /gene="F26F2.4"
 join(10009..10083,10128..10289)
 /gene="F26F2.4"
 /note="Predicted using GeneFINDER"

```
/codon_start=1
/protein_id="CAB04190.1"
/db_xref="GI:3876427"
/db_xref="SPTREMBL:O9YV55"
/translation="MSDNTLMAMEMDKNSQHFQFMAMENNSARLSQARDR
NAIRNENKLEVPNKLEAKKESPIDISIKIR"
join(14010..14084,14132..14293)
/gene="F26F2.5"
join(14010..14084,14132..14293)
/gene="F26F2.5"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB04191.1"
/db_xref="GI:3876428"
/db_xref="SPTREMBL:O9YV54"
/translation="MSSTYQLEMAEMDNQONARONDLTRAMENNSARLSQARDR
NAIRSENNKEVYKIKYKQKREKSPDITSEPR"
join(14918..14971,15019..15312,15358..15433,15482..15852,
15935..16057,16101..16301,16345..16433,16484..16549,
16594..16756,17003..17197,17252..17337,17383..17477,
17840..17958)
/gene="F26F2.6"
join(14918..14971,15019..15312,15358..15433,15482..15852,
15935..16057,16101..16301,16345..16433,16484..16549,
16594..16756,17003..17197,17252..17337,17383..17477,
17840..17958)
/gene="F26F2.6"
/note="predicted using GeneFinder"
/codon_start=1
/protein_id="CAB04192.1"
/db_xref="GI:3876429"
/db_xref="SPTREMBL:O9YV53"
/translation="MKRIYLFCEFTVLLAAQKQSPYKFGVIDEPIIDINSTQVND
FDECEKCEEDNLCILAQKSDSEPCQLEPMNITSGVGENESGYQYAFVYDQPA
CELN.PRLNGKRPYIENNTREYMTIDISEMKISYITRTLDNLRCMGSEFEP
VYDGDANGSVWQVADYAKDGEYTKPDQDISNMECHMGKREBEQOIAVLSQAK
KQWYSSDDLTPTMNNSSADSKRLVTKMYDEPTCLFTTLPOLDSRFQGTQSEPT
TEISTGTFRQSKDPTNARIPAVEGCYKAVRIRIDVLSSDITNNAOQCLEIG
GGILMTVAINDTLNPLKGDYLEDKMTPRHVGIILKDDITWMLPOLKIDQSPLEW
SLMTRKVADEKCGYOLSPDLPFCVPCVPCDDPMYNGYICHPKIDSPDQPT
ATTEMSEYKQSKQADARFIPGVQRCYKMYETKDEPTINAFSHEAAGNCTSG
NGMLVAATNDSMKLFRGILEKHEDEWVGLIKDSNTWMLPLVITDNPIDWAD
OKFNMGDCGYVOLVPRMOKIDEPVRLVSPKRTNNAVKLGSKNOSTVILXSDSE
ILNKNVIRFSLSLIOION"
join(21334..21359,21403..21567,21665..21740,21843..21936,
22373..22469,22827..22992,23123..23234,23506..23612,
24967..25350,26923..27087,27972..28077,28730..28924,
29404..29486,29536..29648,30117..30366)
/gene="F26F2.7"
join(21334..21359,21403..21567,21665..21740,21843..21936,
22373..22469,22827..22992,23123..23234,23506..23612,
24967..25350,26923..27087,27972..28077,28730..28924,
29404..29486,29536..29648,30117..30366)
/note="cDNA EST YK395f6.5 comes from this gene"
/codon_start=1
/protein_id="CAB62802.1"
/db_xref="GI:6562358"
/db_xref="SPTREMBL:O9U3B8"
/translation="MANNVSTIRENSRNPSSSSSIDQLINVEANDLSVYVEDVSI
TMSAPPSPATPRKIDPSAETEPKSGIRRLTIDKISGDLNDPESTPAIRPSRI
SYDERLIVICDSOTSSPLOTSPSOERSDAPAVPOPGFSPFWEELAGYSLAND
HARPSRRKRYAARIPLELEPFLITLLOCTDAGSYFTPLPIPLRPIRGLIARI
KRTMSAETCPFLVYITVAASMLIREIDSSLTIVQHSQGVYKIKITFNMEVADRIF
SSUGODIDALMTANSEKRESVGFITSCGHLVALTALHSEVLVLQATTINAVF
NSHNOYVLAISMNSNVELKGVKFKAKAKRQMACSDVRLRFPALFVWIRMA
TAVNNOTVLAITEMNPIDILVWGVCEYVDMKAKHATIKFENIAEYKFPITFVIRI
RSRDSASVSDYSPQSRMGFIPISLIMIRVSCFTEDMWSGCIIFGGLIFDVA
YKICGVNMLGOACHVKKFRFDIOARAEFEIRKRYEKKSSQANSRMLIDETTYA
LHOPAKGFTVSDMTSCWEELOPSTLSSETRSTRTETRAVSHLTARDEDETPRAS
MAHIDPRDSEPPAPASMDQDQDTEPTVPTENNNSNSBOASVYKKTITAAVYTS
ASTNTNATSSDELADVATKAPDQGVQKITE"
```

BASE COUNT

13643

a 7672 c 7543 g 13497 t

```
ORIGIN
Query Match 100.0%; Score 16; DB 3; Length 42355;
Best Local Similarity 100.0%; Pred. No. 8,7e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTTGTCATTTC 16
|||||
Db 22646 TTGTTGTCATTTC 22661

RESULT 8
AX059515
LOCUS AX059515 48128 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 248 from Patent WO005325.
ACCESSION AX059515
VERSION AX059515.1 GI:12311622
KEYWORDS thale cress.
SOURCE Arabidopsis thaliana
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1 (bases 1 to 48128)
Preuss D., Copenhagen, G. and Keith, K.
Plant chromosome compositions and methods
Patent: WO 005325-A 248 21-SEP-2000;
The University of Chicago (US)
Location/Qualifiers
1..48128
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"

BASE COUNT 15729 a 9023 c 7665 g 15711 t

ORIGIN
Query Match 100.0%; Score 16; DB 6; Length 48128;
Best Local Similarity 100.0%; Pred. No. 8,5e+02;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TTGTTGTCATTTC 16
|||||
Db 38659 TTGTTGTCATTTC 38708

RESULT 9
AC091880
LOCUS AC091880 93711 bp DNA linear PRI 27-FEB-2002
DEFINITION Homo sapiens chromosome 5 clone CTD-3203N7, complete sequence.
ACCESSION AC091880
VERSION AC091880.2 GI:18854960
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 93711)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
2 (bases 1 to 93711)
Unpublished
3 (bases 1 to 93711)
DOE Joint Genome Institute.
Direct Submission
Submitted (09-JUN-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
4 (bases 1 to 93711)
DOE Joint Genome Institute and Stanford Human Genome Center.
Direct Submission
Submitted (21-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
REFERENCE
1 DOE Joint Genome Institute and Stanford Human Genome Center.
AUTHORS
TITLE
JOURNAL
Submitted (27-FEB-2002) DOE Joint Genome Institute, 2800 Mitchell
```


Query Match	100.0%	Score 16;	DB 8;	length 94302;
Best local Similarity	100.0%	Pred. No. 7.7e+02;		
Matches	16;	Conservative	0;	Mismatches 0; Indels 0; Gaps 0;
QY 1	TTGTTGTCATTTTTC 16			
Db	83688 TTGTTGTCATTTTTC 83673			
RESULT 11				
LOCUS	AC093973/c	101445 bp	DNA	linear HTG 11-JUL-2007
DEFINITION	Rattus norvegicus clone CH230-9G14, *** SEQUENCING IN PROGRESS ***			
ACCESSION	AC093973			
VERSION	AC093973.4			
KEYWORDS	HTG; PHASE1.			
SOURCE	Norway rat.			
ORGANISM	Rattus norvegicus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
	1 (bases 1 to 101445)			
	Munry,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C., Alsbrooks,S.L., Amaratunga,H.C., Are,J.R., Ayele,M., Banks,T., Barbacia,J., Beaton,J., Bamage,K., Blankenburg,K., Bonnin,D., Bouck,J., Bowie,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P., Bunyah,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C., Cleeland,C.D., Cox,C., Coyle,M.D., Dethorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Donahue,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frazant,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,D.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlik,P., Hawes,A., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B., Homsi,F., Howard,S., Huber,J., Huik,S., Hume,C., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Kovach,J., Kovar,C., Kratochvic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L., Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Louissege,H., Losado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J., Maheeswari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E., Massey,E., Mawney,E., McLeod,M.P., Meador,M., Mei,G., Metzner,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokwenkwo,S., Ogun,M., Okunodu,G., Orlunge,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Puy,L., Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojibokan,I., Rolfe,M., Ruiz,S., Savary,G., Scherer,S., Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Swalek,A., Tabor,P., Taneitsa,A., Tamezits,K., Tang,H., Tausey,J., Taylor,C., Taylor,T., Telirod,B., Thomas,N., Thomas,S., Umanik,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,			

Williams, G., Williamson, A., Wleczek, R., Woodén, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G., and Gibbs, R.
 TITLE
 JOURNAL
 Unpublished
 2 (bases 1 to 101445)
 REFERENCE
 TITLE
 JOURNAL
 Submitted (13-SEP-2001) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 101445)
 REFERENCE
 TITLE
 JOURNAL
 Submitted (11-JUL-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 On Jul 10, 2002 this sequence version replaced gi:17942322.
 COMMENT

----- Genome Center
 Center: Baylor College of Medicine
 Center code: BCM
 Web site: <http://www.hgsc.bcm.tmc.edu/>
 Contact: hgsc-help@bcm.tmc.edu
 ----- Project Information
 Center project name: GDEJ
 Center clone name: CH230-9G14
 ----- Summary Statistics
 Sequencing vector: Plasmid:
 Chemistry: Dye-terminator Big Dye: 100% of reads
 Assembly program: Phrap: version 0.990329
 Consensus quality: 46253 bases at least Q40
 Consensus quality: 49656 bases at least Q30
 Consensus quality: 52180 bases at least Q20

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 54 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 1198: contig of 1198 bp in length
 * 1299: gap of unknown length
 * 1299: contig of 1320 bp in length
 * 2619: gap of unknown length
 * 2719: contig of 1041 bp in length
 * 3760: gap of unknown length
 * 3859: gap of unknown length
 * 4895: contig of 1036 bp in length
 * 4896: gap of unknown length
 * 4895: gap of unknown length
 * 6090: gap of unknown length
 * 6090: contig of 1094 bp in length
 * 6190: gap of unknown length
 * 7119: contig of 1530 bp in length
 * 7720: gap of unknown length
 * 7820: contig of 1241 bp in length
 * 9060: contig of 1229 bp in length
 * 9061: gap of unknown length
 * 10389: contig of 1229 bp in length
 * 10390: gap of unknown length
 * 10489: gap of unknown length
 * 10490: contig of 1009 bp in length
 * 11498: gap of unknown length
 * 11499: gap of unknown length
 * 11599: contig of 1092 bp in length
 * 12691: gap of unknown length
 * 12791: contig of 1737 bp in length
 * 14528: gap of unknown length
 * 14627: gap of unknown length
 * 14628: contig of 1087 bp in length
 * 15715: gap of unknown length
 * 15715: contig of 1037 bp in length
 * 15815: gap of unknown length
 * 16852: gap of unknown length
 * 16952: contig of 1196 bp in length
 * 18148: gap of unknown length
 * 18247: gap of unknown length
 * 18248: contig of 1442 bp in length
 * 19689: contig of 1442 bp in length

19690: gap of unknown length
 19790: contig of 1025 bp in length
 20815: gap of unknown length
 20914: gap of unknown length
 20915: contig of 1138 bp in length
 22052: gap of unknown length
 22053: gap of unknown length
 22152: contig of 1152 bp in length
 22153: gap of unknown length
 23405: gap of unknown length
 23405: contig of 1707 bp in length
 25111: gap of unknown length
 25112: gap of unknown length
 25112: contig of 1362 bp in length
 26573: gap of unknown length
 26574: gap of unknown length
 27730: contig of 1057 bp in length
 27731: gap of unknown length
 27830: gap of unknown length
 29263: contig of 1433 bp in length
 29264: gap of unknown length
 29264: gap of unknown length
 29364: contig of 1019 bp in length
 30383: gap of unknown length
 30482: gap of unknown length
 32152: contig of 1670 bp in length
 32153: gap of unknown length
 32252: gap of unknown length
 33413: contig of 1161 bp in length
 33413: gap of unknown length
 33414: gap of unknown length
 33414: contig of 1208 bp in length
 34722: gap of unknown length
 34821: gap of unknown length
 36652: contig of 1831 bp in length
 36752: gap of unknown length
 36753: gap of unknown length
 38231: contig of 1479 bp in length
 38232: gap of unknown length
 38332: gap of unknown length
 38332: contig of 1257 bp in length
 39589: gap of unknown length
 39688: gap of unknown length
 42426: contig of 2738 bp in length
 42426: gap of unknown length
 42527: gap of unknown length
 42527: contig of 1747 bp in length
 44273: gap of unknown length
 44274: gap of unknown length
 44374: contig of 1640 bp in length
 46014: gap of unknown length
 46114: gap of unknown length
 47379: contig of 1266 bp in length
 47380: gap of unknown length
 47480: gap of unknown length
 49103: contig of 1623 bp in length
 49202: gap of unknown length
 49203: gap of unknown length
 50462: contig of 1260 bp in length
 50463: gap of unknown length
 50562: gap of unknown length
 50653: contig of 1602 bp in length
 52165: gap of unknown length
 52165: contig of 1421 bp in length
 52265: gap of unknown length
 53685: gap of unknown length
 53685: contig of 1333 bp in length
 53786: gap of unknown length
 55118: gap of unknown length
 55218: gap of unknown length
 55219: contig of 1061 bp in length
 56279: gap of unknown length
 56380: gap of unknown length
 56380: contig of 1638 bp in length
 58017: gap of unknown length
 58017: gap of unknown length
 58117: gap of unknown length
 60356: contig of 2239 bp in length
 60456: gap of unknown length
 60357: gap of unknown length
 63214: contig of 2758 bp in length
 60457: gap of unknown length
 63314: gap of unknown length
 63315: gap of unknown length
 63315: contig of 2508 bp in length
 65822: gap of unknown length
 65822: gap of unknown length
 65823: gap of unknown length
 65923: contig of 2606 bp in length
 68528: gap of unknown length
 68629: gap of unknown length
 71060: contig of 2432 bp in length
 71061: gap of unknown length
 73523: contig of 2363 bp in length
 73524: gap of unknown length
 73524: gap of unknown length
 73624: contig of 2018 bp in length
 75641: gap of unknown length
 75642: gap of unknown length
 79589: contig of 3848 bp in length
 79590: gap of unknown length
 79590: gap of unknown length
 81627: contig of 1938 bp in length
 81628: gap of unknown length
 81628: gap of unknown length
 85510: contig of 3783 bp in length
 85511: gap of unknown length
 85511: gap of unknown length
 89244: contig of 3634 bp in length
 89245: gap of unknown length

* 89345 93002: contig of 3658 bp in length
 * 93003 93102: gap of unknown length
 * 93103 93174: contig of 4072 bp in length
 * 97175 97274: gap of unknown length
 * 97275 101445: contig of 4171 bp in length.

Query Match 100.0%; Score 16; DB 2; Length 101445;
 Best Local Similarity 100.0%; Pred. No. 7.6e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTCATTTTC 16
 |||||||
 Db 63923 TTGTGTCATTTTC 63908

RESULT 12
 ALL57407/C
 LOCUS
 DEFINITION Human DNA sequence from clone RP4-609e1 on chromosome 1p31.2-32.1,
 complete sequence.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

ALL57407 142396 bp DNA linear PRI 15-NOV-2001
 Human DNA sequence from clone RP4-609e1 on chromosome 1p31.2-32.1,
 complete sequence.
 ALL57407
 ALL57407.11 GI:16972784
 HTG.
 human.
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 1 (bases 1 to 142396)
 Leongamornlert, D.
 Direct Submission
 Submitted (15-NOV-2001) Wellcome Trust Sanger Institute, Hinxton,
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:
 humanyes@sanger.ac.uk Clone requests: clonerequests@sanger.ac.uk
 On Nov 16, 2001 this sequence version replaced gi:1590765.
 During sequence assembly data is compared from overlapping clones.
 Where differences are found these are annotated as variations
 together with a note of the overlapping clone name. Note that the
 variation annotation may not be found in the sequence submission
 corresponding to the overlapping clone, as we submit sequences with
 only a small overlap as described above.
 This sequence was finished as follows unless otherwise noted: all
 regions were either double-stranded or sequenced with an alternate
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such
 as compressions and repeats: all regions were covered by at least
 one plasmid subclone or more than one M13 subclone; and the
 assembly was confirmed by restriction digest. The following
 abbreviations are used to associate primary accession numbers given
 in the feature table with their source databases: Em: EMBL; Sw:
 SWISSPROT; Tr: TrEMBL; Wp: WormPeP; Information on the WormPeP
 database can be found at
 http://www.sanger.ac.uk/projects/C_elegans/wormpep This sequence
 was generated from part of bacterial clone contigs of human
 chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping
 Group. Further information can be found at
 http://www.sanger.ac.uk/HGP/Chr1
 RP4-609e1 is from the library RPc1-4 constructed by the group of
 Pieter de Jong. For further details see
 http://www.chori.org/bacpac/home.htm
 VECTOR: pCYPAC2

IMPORTANT: This sequence is not the entire insert of clone
 RP4-609e1. It may be shorter because we sequence overlapping
 sections only once, except for a short overlap.
 The true left end of clone RP4-609e1 is at 1 in this sequence. The
 true left end of clone RP1-518D3 is at 140397 in this sequence.
 The true right end of clone RP11-422P8 is at 3481 in this sequence.

FEATURES
 source
 1. 142396
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /map="p31.2-32.1"
 /clone="RP4-609e1"

BASE COUNT 40657 a 28410 c 27397 g 43932 t
 ORIGIN

Query Match 100.0%; Score 16; DB 9; Length 142396;
 Best Local Similarity 100.0%; Pred. No. 7.2e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTCATTTTC 16
 |||||||
 Db 97887 TTGTGTCATTTTC 97872

RESULT 13
 AC105683/C
 LOCUS
 DEFINITION Rattus norvegicus clone CH230-268K18, *** SEQUENCING IN PROGRESS
 *** 45 unordered pieces.
 ACCESSION
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM

REFERENCE
 AUTHORS
 1 (bases 1 to 149834)

Muzny, D.M., Adams, C., Adlo-Oduola, B., Ali-osman, F.R., Allen, C.,
 Alsbrooks, S.L., Amarantunga, H.C., Are, J.R., Ayele, M., Banks, T.,
 Barbarta, J., Benton, J., Blum, E., Brown, E., Brown, N.P., Bryant, N.P.,
 Buhay, C., Burck, P., Burkett, C., Burrell, K.L., Byrd, N.C.,
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,
 Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopoulos, C.,
 Cleveland, C.D., Cox, C., Coyle, M.D., Dalbourn, S.R., David, R.,
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,
 Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,
 Eamhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escoto, M.,
 Falls, T., Ferriguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,
 Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,
 Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,
 Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,
 Hernandez, O., Hodgson, A., Hognes, M., Holloway, C., Hollins, L.E.,
 Honsi, F., Howard, S., Huber, J., Hulik, S., Hume, J., Jackson, L.E.,
 Jacobson, B., Jia, T., Johnson, R., Joly, S., Joudan, S.,
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,
 Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L.,
 Li, J., Li, Z., Lichtarge, O., Ilen, C., Liu, J., Liu, N., Louisedge, H.,
 Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,
 Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,
 Massey, E., Maminney, E., McLeod, K.P., Meador, M., Mei, G., Metzger, M.,
 Miner, G., Miner, Z., Mitchell, T., Mohabhat, K., Morgan, M., Morris, S.,
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,
 Nguyen, N., Nickerson, E., Nwokkenko, S., Ogih, M., Okunou, G.,
 Orangunye, N., Oriedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,
 Peters, L., Pickens, R., Prins, E., Pu, L.L., Quiles, M., Ren, Y.,
 Rives, M., Rojas, A., Rojudoan, I., Rolfe, M., Ruiz, S., Savery, G.,
 Scherer, S., Scott, G., Shen, H., Shoshitari, N., Sisson, I.,
 Sodergren, E., Sonaike, T., Sparks, A., Stanley, K., Stone, H.,
 Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, O.,
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wallington, S.,
 Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,
 Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
 Weinstein, G. and Gibbs, R.

TITLE
 JOURNAL
 REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 2 (bases 1 to 149834)
 Morley, K.C.
 Direct Submission
 Submitted (09-JAN-2002) Human Genome Sequencing Center, Department

REFERENCE
AUTHORS
TITLE
JOURNAL

of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 149834)

Worley/K.C.
Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

On Jul 12, 2002 this sequence version replaced g1:18092906.

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

Project Information

Center project name: GNMU

Center clone name: CH230-268K18

Summary Statistics

Sequencing vector: Plasmid

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 112644 bases at least Q40

Consensus quality: 118003 bases at least Q30

Consensus quality: 122664 bases at least Q20

* NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 45 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```

1 1206: contig of 1206 bp in length
* 1207 1306: gap of unknown length
* 1307 2771: contig of 1465 bp in length
* 2772 2871: gap of unknown length
* 2872 4291: contig of 1420 bp in length
* 4292 4391: gap of unknown length
* 4392 5890: contig of 1499 bp in length
* 5891 5990: gap of unknown length
* 5991 7003: contig of 1013 bp in length
* 7004 7104: gap of unknown length
* 7104 8760: contig of 1657 bp in length
* 8761 8860: gap of unknown length
* 8861 10203: contig of 1343 bp in length
* 10204 10304: gap of unknown length
* 10304 12057: contig of 1753 bp in length
* 12057 12157: gap of unknown length
* 12157 14304: contig of 2148 bp in length
* 14305 14404: gap of unknown length
* 14405 15435: contig of 1031 bp in length
* 15436 15535: gap of unknown length
* 15536 17038: contig of 1503 bp in length
* 17039 17138: gap of unknown length
* 17139 18996: contig of 1858 bp in length
* 18997 19097: gap of unknown length
* 19097 20300: contig of 1204 bp in length
* 20301 20400: gap of unknown length
* 20401 22309: contig of 1909 bp in length
* 22310 22409: gap of unknown length
* 22410 24216: contig of 1807 bp in length
* 24217 24316: gap of unknown length
* 24317 25433: contig of 1117 bp in length
* 25434 25533: gap of unknown length
* 25534 27232: contig of 1699 bp in length
* 27233 27332: gap of unknown length
* 27333 28906: contig of 1574 bp in length
* 28907 31205: gap of unknown length
* 31206 31305: contig of 2199 bp in length
* 31305: gap of unknown length

```

```

* 31306 33683: contig of 2378 bp in length
* 33684 33783: gap of unknown length
* 33784 36084: contig of 2301 bp in length
* 36085 36184: gap of unknown length
* 36185 38010: contig of 1826 bp in length
* 38011 38110: gap of unknown length
* 38111 40318: contig of 2208 bp in length
* 40319 40418: gap of unknown length
* 40419 43000: contig of 2582 bp in length
* 43001 43100: gap of unknown length
* 43101 45777: contig of 2677 bp in length
* 45778 45877: gap of unknown length
* 45878 47683: contig of 1806 bp in length
* 47684 47783: gap of unknown length
* 47784 50120: contig of 2337 bp in length
* 50121 50220: gap of unknown length
* 50221 53769: contig of 3549 bp in length
* 53770 53869: gap of unknown length
* 53870 56450: contig of 2581 bp in length
* 56451 56550: gap of unknown length
* 56551 59806: contig of 3256 bp in length
* 59807 59906: gap of unknown length
* 59907 62490: contig of 2584 bp in length
* 62491 62590: gap of unknown length
* 62591 66273: contig of 3683 bp in length
* 66274 66373: gap of unknown length
* 66374 70744: contig of 4371 bp in length
* 70745 70844: gap of unknown length
* 70845 73643: contig of 2799 bp in length
* 73644 73743: gap of unknown length
* 73744 78206: contig of 4463 bp in length
* 78207 78306: gap of unknown length
* 78307 82599: contig of 4293 bp in length
* 82600 82699: gap of unknown length
* 82700 88000: contig of 5301 bp in length
* 88001 88100: gap of unknown length
* 88101 93518: contig of 5418 bp in length
* 93519 93618: gap of unknown length
* 93619 99466: contig of 5848 bp in length
* 99467 99566: gap of unknown length
* 99567 106975: contig of 7409 bp in length
* 106976 107076: gap of unknown length
* 107077 113718: contig of 6643 bp in length
* 113719 113818: gap of unknown length
* 113819 120398: contig of 6580 bp in length
* 120399 120498: gap of unknown length
* 120499 129737: contig of 9239 bp in length
* 129738 129837: gap of unknown length
* 129838 140296: contig of 10459 bp in length
* 140297 140396: gap of unknown length
* 140397 149834: contig of 9438 bp in length.

```

FEATURES

Location/Qualifiers

1..149834

source

"/organism="Rattus norvegicus"

BASE COUNT 43738 a 28959 c 29383 g 43046 t 4708 others

ORIGIN

Query Match 100.0%; Score 16; DB 2; Length 149834;

Best Local Similarity 100.0%; Pred. No. 7; len=02;

Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGTGTCAATTTC 16

Db 19945 TTGTGTCAATTTC 19930

RESULT 14

AC117244 161658 bp DNA linear HTG 03-MAY-2002

LOCUS AC117244 MUS MUSCULUS CHROMOSOME UNK CLONE RP24-22SD10, WORKING DRAFT

DEFINITION SEQUENCE, 4 unordered pieces.

ACCESSION AC117244
 VERSION AC117244.2 GI:20429657
 HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
 KEYWORDS Mus musculus.
 SOURCE Mus musculus.
 ORGANISM Mus musculus.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 161658)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE The sequence of Mus musculus clone
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 161658)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (08-APR-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 REFERENCE 3 (bases 1 to 161658)
 AUTHORS McPherson, J.D. and Waterston, R.H.
 TITLE Direct Submission
 JOURNAL Submitted (03-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
 COMMENT On May 3, 2002 this sequence version replaced gi:20069824.

----- Genome Center -----
 Center: Washington University Genome Sequencing Center
 Center code: WUGSC
 Web site: http://genome.wustl.edu/gsc/index.shtml
 Contact: submissions@wustl.edu
 Project Information
 Center project name: M_BB0225D10
 ----- Summary Statistics -----
 Sequencing vector: M13; 0%
 Sequencing vector: Plasmid; 100%
 Chemistry: Dye-primer ET; 0% of reads
 Chemistry: Dye-terminator Big Dye; 100% of reads
 Assembly program: Phrap; version 0.990319
 Consensus quality: 160286 bases at least Q40
 Consensus quality: 160767 bases at least Q30
 Consensus quality: 161015 bases at least Q20
 Insert size: 169000; agarose-fp
 Insert size: 161358; sum-of-contigs
 Quality coverage: 10.48 in Q20 bases; agarose-fp
 Quality coverage: 10.21 in Q20 bases; sum-of-contigs
 ----- NOTE: This is a 'working draft' sequence. It currently consists of 4 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as * runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved. -----
 1 1453: contig of 1453 bp in length
 * 1454 1553: gap of unknown length
 * 1554 2783: contig of 1230 bp in length
 * 2784 2883: gap of unknown length
 * 2884 8051: contig of 5168 bp in length
 * 8052 8152: gap of unknown length
 * 8152 161658: contig of 153507 bp in length.
 FEATURES
 source
 1. 161658
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /chromosome="X"
 /clone="RP24-225D10"
 1. 1453
 /note="assembly_name:Contig23"
 misc_feature
 1554. 2783
 /note="assembly_name:Contig24"
 misc_feature
 2884. 8051
 /note="assembly_name:Contig25"
 misc_feature
 8152. 161658

BASE COUNT 51109 a 31750 c 31043 g 47456 t 300 others
 ORIGIN
 Query Match 100.0%; Score 16; DB 2; Length 161658;
 Best Local Similarity 100.0%; Pred. No. 7.1e+02;
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 0Y 1 TTGTGTCAATTTTTC 16
 |||||||
 DB 144084 TTGTGTCAATTTTTC 144099
 RESULT 15
 AC016256
 LOCUS
 DEFINITION
 AC016256 Homo sapiens 12 BAC Rpl1-98B5 (Roswell Park Cancer Institute Human BAC library) complete sequence.
 VERSION
 AC016256.19 GI:14670071
 KEYWORDS
 AC016256.19 GI:14670071
 SOURCE
 ORGANISM
 Homo sapiens.
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 1 (bases 1 to 164405)
 AUTHORS
 Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-Osman, F.R., Allen, C., Alstbrooks, S.L., Amaralunge, H.C., Are, J.R., Banks, T., Barbara, J., Benton, J., Blum, K., Blankenburg, K., Bonnin, D., Bouck, J., Bowie, S., Brieva, M., Brown, E., Brown, M., Bryant, N.P., Buhay, C., Burich, P., Burkett, C., Burrell, K.L., Byrd, N.C., Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Z., Chiu, D., Chowdhry, I., Christopoulos, C., Cleveland, C.D., Cox, C., Coyle, M.D., Dachtone, S.R., David, R., Davila, M.L., Davis, C., Day, Carroll, L., Dederich, D.A., Delaney, K.J., Draper, H., Dunn, A.L., Ding, Y., Dinh, H.D., Douthett, C., Edgar, D., Edwards, C.C., Dugan-Rocha, S., Durbin, K.J., Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Emerling, S., Escotto, M., Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P., Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R., Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K., Han, J., Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J., Hernandez, O., Hodgson, A., Hogue, M., Hollway, C., Hollins, B., Homsi, F., Howard, S., Huber, J., Hui, S., Hume, J., Ioshikhes, I., Jackson, L.E., Jacobson, B., Jia, Y., Johnson, R., Jolyet, S., Joudah, S., Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C., Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lee, E., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Liu, C., Liu, D., Liu, W., Louisedge, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapa, P., Marondel, I., Martin, R., Martindale, A., Martinez, E., Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Mercher, S., Metzker, M., Miller, A., Miner, G., Miner, Z., Mitchell, T., Mohabdal, K., Montgomery, K.T., Morgan, M., Morris, S., Moser, M., Neal, D., Nelson, D., Newton, E., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokkenko, S., Ogwu, M., Okunolu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Pilius, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojoudan, I., Rolfe, M., Ruiz, S., Saverly, G., Scherer, S., Scott, G., Shen, H., Shim, C., Shoostali, N., Sisson, I., Sodergren, E., Sonak, T., Sparks, A., Stanley, H., Stone, H., Sutton, A., Sytek, A., Taylor, P., Tamersia, A., Tamerisa, K., Tang, H., Tansey, J., Taylor, C., Taylor, B., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczky, R., Woodson, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Zuchelapatti, R. and Gibbs, R.
 TITLE
 JOURNAL
 AC016256
 AUTHORS
 TITLE
 JOURNAL
 Direct Submission
 2 (bases 1 to 164405)
 Worley, K.C.
 Direct Submission
 Submitted (24-NOV-1999) Human Genome Sequencing Center, Department

Fri Jan 24 10:08:59 2003

us-09-803-736-1022_copy_10615_10630.rge

Page 16

OY 1 TTGTGTCATTTC 16
|||||
Db 4507 TTGTGTCATTTC 4522

Search completed: January 23, 2003, 21:22:18
Job time : 1312.06 secs

=> fil reg; d que 16

FILE 'REGISTRY' ENTERED AT 15:16:55 ON 24 JAN 2003

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2003 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file provided by InfoChem.

STRUCTURE FILE UPDATES: 23 JAN 2003 HIGHEST RN 480990-41-8

DICTIONARY FILE UPDATES: 23 JAN 2003 HIGHEST RN 480990-41-8

TSCA INFORMATION NOW CURRENT THROUGH MAY 20, 2002

Please note that search-term pricing does apply when conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. See HELP PROPERTIES for more information. See STNote 27, Searching Properties in the CAS Registry File, for complete details:
<http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf>

L6 3 SEA FILE=REGISTRY ABB=ON UGAGUGUAACUCUCCCCGCAAGACUAAGGAAUCUGCC
CUUCCAACCCAGCCAGCCUUGUUGUCAWUUUUUCCAAGAUUACCCGAAAGUGUCCUUAUUAAU
UCUCCUCUGAAGGACCGGCA|UGCCGGUCCUUCAGAGGAGAAUUAUAAGGACACUUUCGGUG
AUAUCUUGGAAAAAWUGACAACAAGGCUGGCUGGUUGGAAGGGCAGAUUCCUAGUCUUGCGG
GGAGAGUUACACUCA/QSN

=> d rn cn sql kwic nte lc 16 1-3

L6 ANSWER 1 OF 3 REGISTRY COPYRIGHT 2003 ACS

RN 316120-05-5 REGISTRY

CN GenBank AX059515 (9CI) (CA INDEX NAME)

SQL 48128

SEQ 38601 catggtgtga accggaatcg acgaaagcac tgccttggtg agtgtaactc

38651 tccccgcaag actaaggaat ctgcccttcc aaccagccag ccttggtgtc

38701 aattttttcca agatatcacc gaaagtgtcc ttattaattc tcctctgaag

38751 gaccggcatt ccaaggtatt ttccaagctc tctagttgag ctaatccac

HITS AT: 38639-38758

NTE doublestranded

LC STN Files: GENBANK

L6 ANSWER 2 OF 3 REGISTRY COPYRIGHT 2003 ACS

RN 295809-41-5 REGISTRY

CN DNA (Arabidopsis thaliana chromosome 2 contig_1 centromere-containing fragment) (9CI) (CA INDEX NAME)

OTHER NAMES:

CN 4: PN: WO0055325 SEQID: 209 claimed DNA

SQL 611549

SEQ 228101 aaccggaatc gacgaaagca ctgcccttggg gagtgtaact ctccccgcaa

228151 gactaaggaa tctgcccttc caaccagcca gccttggtgt caatttttcc

228201 aagatatcac cgaaagtgtc cttattaatt ctctcttgaa ggaccggcat
=====

HITS AT: 228130-228249

LC STN Files: CA, CAPLUS, TOXCENTER

L6 ANSWER 3 OF 3 REGISTRY COPYRIGHT 2003 ACS

RN 225776-51-2 REGISTRY

CN DNA (Arabidopsis thaliana chromosome 2 section 44 of 255) (9CI) (CA INDEX
NAME)

OTHER NAMES:

CN GenBank AC007730

CN GenBank AE002093

SQL 94302

SEQ 83601 gagcttgaa aataccttgg aatgccggtc cttcagagga gaattaataa
=====

83651 ggacactttc ggtgatattc tggaaaaatt gacaacaagg ctggctggtt
=====

83701 ggaagggcag attccttagt cttgcgggga gagttacact caccaaggca
=====

HITS AT: 83623-83742

LC STN Files: CA, CAPLUS, GENBANK

=> fil capl ; s 16; fil toxcenter; s 16

FILE 'CAPLUS' ENTERED AT 15:17:35 ON 24 JAN 2003

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

Copyright of the articles to which records in this database refer is held by the publishers listed in the PUBLISHER (PB) field (available for records published or updated in Chemical Abstracts after December 26, 1996), unless otherwise indicated in the original publications. The CA Lexicon is the copyrighted intellectual property of the American Chemical Society and is provided to assist you in searching databases on STN. Any dissemination, distribution, copying, or storing of this information, without the prior written consent of CAS, is strictly prohibited.

FILE COVERS 1907 - 24 Jan 2003 VOL 138 ISS 5

FILE LAST UPDATED: 23 Jan 2003 (20030123/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

L8 2 L6

FILE 'TOXCENTER' ENTERED AT 15:17:45 ON 24 JAN 2003

COPYRIGHT (C) 2003 ACS

FILE COVERS 1907 TO 20 Jan 2003 (20030120/ED)

This file contains CAS Registry Numbers for easy and accurate substance identification.

TOXCENTER has been enhanced with new files segments and search fields. See HELP CONTENT for more information.

TOXCENTER thesauri in the /CN, /CT, and /MN fields incorporate the MeSH 2003 vocabulary. See <http://www.nlm.nih.gov/mesh/summ2003.html> for a description on changes.

L9 1 L6

=> dup rem 18 19

FILE 'CAPLUS' ENTERED AT 15:17:52 ON 24 JAN 2003
USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.
PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2003 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'TOXCENTER' ENTERED AT 15:17:52 ON 24 JAN 2003

COPYRIGHT (C) 2003 ACS

PROCESSING COMPLETED FOR L8

PROCESSING COMPLETED FOR L9

L10 2 DUP REM L8 L9 (1 DUPLICATE REMOVED)
ANSWERS '1-2' FROM FILE CAPLUS

=> d ibib ab hitrn l10 1-2

L10 ANSWER 1 OF 2 CAPLUS COPYRIGHT 2003 ACS DUPLICATE 1
ACCESSION NUMBER: 2000:666878 CAPLUS
DOCUMENT NUMBER: 133:262304
TITLE: Arabidopsis thaliana chromosome centromere sequences
and their use in DNA constructs and vectors
INVENTOR(S): Preuss, Daphne; Copenhaver, Gregory; Keith, Kevin
PATENT ASSIGNEE(S): The University of Chicago, USA
SOURCE: PCT Int. Appl., 1451 pp.
CODEN: PIXXD2
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 2
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
WO 2000055325	A2	20000921	WO 2000-US7392	20000317
W:	AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW, AM, AZ, BY, KG, KZ, MD, RU, TJ, TM			
RW:	GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW, AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG			
BR 2000009119	A	20011226	BR 2000-9119	20000317
EP 1165792	A2	20020102	EP 2000-916559	20000317
R:	AT, BE, CH, DE, DK, ES, FR, GB, GR, IT, LI, LU, NL, MC, IE, FI			
PRIORITY APPLN. INFO.:			US 1999-125219P	P 19990318
			US 1999-127409P	P 19990401
			US 1999-134770P	P 19990518
			US 1999-153584P	P 19990913
			US 1999-154603P	P 19990917
			US 1999-172493P	P 19991216
			WO 2000-US7392	W 20000317
AB	The present invention provides for the mapping via tetrad anal., cloning, and sequencing of functional plant centromeres from Arabidopsis thaliana chromosomes 1-5. This will permit construction of stably inherited minichromosomes which can serve as vectors for the construction of			

transgenic plant and animal cells. In addn., information on the structure and function of these regions will prove valuable in isolating addnl. centrometric and centromere-related genetic elements and polypeptides from other species.

IT 295809-41-5

RL: BOC (Biological occurrence); BSU (Biological study, unclassified); BUU (Biological use, unclassified); PRP (Properties); BIOL (Biological study); OCCU (Occurrence); USES (Uses)

(nucleotide sequence; Arabidopsis thaliana chromosome centromere sequences and their use in DNA constructs and vectors)

L10 ANSWER 2 OF 2 CAPLUS COPYRIGHT 2003 ACS

ACCESSION NUMBER: 2000:9181 CAPLUS

DOCUMENT NUMBER: 132:89085

TITLE: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana

AUTHOR(S): Lin, Xiaoying; Kaul, Samir; Rounsley, Steve; Shea, Terrance P.; Benito, Maria-Lnes; Town, Christopher D.; Fujii, Claire Y.; Mason, Tanya; Bowman, Cheryl L.; Barnstead, Mary; Feldblyum, Tamara V.; Buell, C. Robin; Ketchum, Karen A.; Lee, John; Ronning, Catherine M.; Koo, Hean L.; Moffat, Kelly S.; Cronin, Lisa A.; Shen, Mian; Pai, Grace; Van Aken, Susan; Umayam, Lowell; Tallon, Luke J.; Gill, John E.; Adams, Mark D.; Carrera, Ana J.; Creasy, Todd H.; Goodman, Howard M.; Somerville, Chris R.; Copenhaver, Greg P.; Preuss, Daphne; Nierman, William C.; White, Owen; Eisen, Jonathan A.; Salzberg, Steven L.; Fraser, Claire M.; Venter, J. Craig

CORPORATE SOURCE: The Institute for Genomic Research, Rockville, MD, 20850, USA

SOURCE: Nature (London) (1999), 402(6763), 760-768
CODEN: NATUAS; ISSN: 0028-0836

PUBLISHER: Macmillan Magazines

DOCUMENT TYPE: Journal

LANGUAGE: English

AB Arabidopsis thaliana (Arabidopsis) is unique among plant model organisms in having a small genome (130-140 Mb), excellent phys. and genetic maps, and little repetitive DNA. The sequence of chromosome 2 from the Columbia ecotype is reported in two gap-free assemblies (contigs) of 3.6 and 16 megabases (Mb). The latter represents the longest published stretch of uninterrupted DNA sequence assembled from any organism to date. Chromosome 2 represents 15% of the genome and encodes 4037 genes, 49% of which have no predicted function. Roughly 250 tandem gene duplications were found in addn. to large-scale duplications of about 0.5 and 4.5 Mb between chromosomes 2 and 1 and between chromosomes 2 and 4, resp. Sequencing of nearly 2 Mb within the genetically defined centromere revealed a low d. of recognizable genes, and a high d. and diverse range of vestigial and presumably inactive mobile elements. More unexpected is what appears to be a recent insertion of a continuous stretch of 75% of the mitochondrial genome into chromosome 2.

IT 225776-51-2

RL: BSU (Biological study, unclassified); PRP (Properties); BIOL (Biological study)

(nucleotide sequence; sequence and anal. of chromosome 2 of the plant Arabidopsis thaliana)

REFERENCE COUNT: 23 THERE ARE 23 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT



PubMed	Nucleotide	Protein	Genome	Structure	PMC	Taxonomy	OMIM	Book
Search	Nucleotide	for				Go	Clear	
Limits		Preview/Index		History		Clipboard		Details
Display	default	Show	20	Send to	File	Get Subsequence		

1: AX059515. Sequence 248 from...[gi:12311622]

Links

LOCUS AX059515 48128 bp DNA linear PAT 17-JAN-2001
DEFINITION Sequence 248 from Patent WO0055325.
ACCESSION AX059515
VERSION AX059515.1 GI:12311622
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Rosidae; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
REFERENCE 1
AUTHORS Preuss, D., Copenhaver, G. and Keith, K.
TITLE Plant chromosome compositions and methods
JOURNAL Patent: WO 0055325-A 248 21-SEP-2000;
The University of Chicago (US)
FEATURES Location/Qualifiers
source 1..48128
/organism="Arabidopsis thaliana"
/db_xref="taxon:3702"
BASE COUNT 15729 a 9023 c 7665 g 15711 t
ORIGIN

```
1 aaatcagtcg tgaacaaaac agtaactcgg ccatcaaatac gaatagatct aacaaatcag
61 ccacgtatg taaatcagtg acggcgaaaa tcaaccgtaa taattgtcga taaaaaccct
121 caatatcaca gttttcgtaa aacgggtttca gtcacgggaa tcacttgctg ttgaacttag
181 atgtcgaaaa aatctaatta aagttagttt agacacttac cggaacaagt tcaaatgcga
241 ctgtataggg atgtgacttt actcgactta aaattcgaca ctatcgaagg attttggcgg
301 aaagagatga gataaataga aaacattgac aaaattaggg ttccaattaa ggaatcaagt
361 cgacggaaca agagaagaga gagccagaga atagttcttt tttttttttt tttttttttt
421 gcacgcggtat taaccgtgta gagggaaaaa ggaaaactaa aagggttaac gaaattttcg
481 ggtaatgacg tggattagga taatgacgtg gataatgatt gtttttaaat taaaaaaaaa
541 atattatttta gataaactaa tcaaaggtgt taattgtaat tatacaaatg ctcaaaacat
601 atgagtaatt ttttaaaact ctatgatatt ctagtataaa acatactttt ctatgatatac
661 tgagtaatta attcaccocat ttttaatatata aaaataccaa aaaattagtt ttagattata
721 attcataaaa cctaataagt ttttagttta aaactaatat atcatagtca caactcatga
781 ctaatatgaa ctactgagga acaatgggag attagatcgc atagagttca attcccggtc
841 tcaacccatg accaatatga actcaaagtt tcttcaaaa cctcgttgcc tccgttcgat
901 gcctaattta atagggaaac tcaaagcggc tctatttcat tctattccat acatatccca
961 attccattca ttttaatatca ttttggtgtc attgacataa gagatgtaat ttctagtcta
1021 tctctttata tttatatata tgaaaagttg caaaatcatc atataataat ctagaaattg
1081 aaatagaaaa gaaaaaagag aggtttgtga tggaaatttca atcttttata gtaggtaatac
1141 tagtatcctt atgcatgaag ataataaatt tggtcgttgt ggtcggactc tattgtggat
1201 ttatgaccac attctacata gggccttctt atctcttctt tctccgagct cgggtttatg
1261 acgaagagga agaaggaaac gagaagaaag taccagcaag aacatgtttt attacggggac
1321 agctcatgat gttcatatca atctattatg cgctatgca tttagcattg tgtagacctt
1381 atacaataac tgttctagct ctaccgtatc ttttgtttca tttattatgg aacaatcaca
1441 aacatttttt ttattatgga tctactacca gaaataaaat gtgtaatctt cgcattcaat
1501 gtgtattcct gaataatcgc atattttcaat tattcaacca tttcatttta ccaagttcat
1561 tgttagttgt ggaagccccc ttgcccattg gttgactaag ggttcattaa gggttttaca
1621 ccaagaagtc tgggttcgag tcttagatac tgcgggttat tagcaaaatt cgagattaat
1681 ggagacaagt gtgcaggaaa tctccaacga tgtacaatta tagcggttca ccaggattct
1741 acacagataa gatcattgaa ggtgtccatc aagagcatgg agaaagagaa gaaggtgtcg
```

```
1801 tctgtcgtgg atctaataag ggaggaagga attgcccgta gtacaatcgt ctatgtaaca
1861 gttttcataa tttgcaataa ataataaatc taagacggca aaaataagtt taatgttagc
1921 tagattagtc aacatttata tgtttcgata caacaacaag atgttatttg taacaagtag
1981 tttcgttggg tggttaattg gtcacatttt attcctgcaa agggtttggg ttggtattag
2041 tctcgataca acaaaataat tctattaggt ctaatgtagt tattagatct aataagata
2101 agttccttgg gtcagaattg agaaaattcta tggctcgaat ctttagtatt ctcttactta
2161 ttatctgtgt ctactagtta ggcagaatac catcacccat ttttactaag aaacttaaa
2221 gaacctcaga aaggggtgtg actaaacagg agcaagaggt atccattgaa gaaaccctt
2281 ttccttctat ttttttgaag gaaggggaag atctggacaa aatcgataaa atgaaagaaa
2341 tcctagttaa tggaaaagac aaaattaata aagatgatga attacacgtt cgaacatact
2401 ataactataa aactgtttct gaaaaccaat atggaaataa agaaaattcg aatttagatt
2461 ttttcaaaaa aacaaaaaag gatcattaaa aaaatcaata cataggacaa ctataagaac
2521 atataagaag aaatgcgact tccacctata tttttgttta ctctcctac aaaatatttt
2581 ttaataccta ctccatttgg aattccatca atgattcttt tatcaaaaaa agtcgtttgt
2641 tttgctaata ttcttatact ttcaattaaa gatgttttta aaaaaagaat ctatgtaacc
2701 acgattatat gaccaattat atacaaaaat taatagtttt tcccaatgaa tttgtttagg
2761 atttcatttt ggaaatgaac ttagtaaggt taaatttaat aaagatgaat aaaaaggctt
2821 atataaacag tatgtctata atattccaaa aaggctatat atattgaccg aaaaagttgc
2881 atttctcaaa agttcatacc aatctacaaa attttgtgaa tttttatgca aaaggtttat
2941 tgacggcgtg aaatattttg ataatatatc aaagtctatt cttttttgat ttaaaggaat
3001 tcctatgggt ccaataaaaa ggtaataaaa accaaaaataa gcataggaaa tagaacagta
3061 ttgtctgatt catgcggtata atagaaagtt cttgtattaa gtccacaaag agagagtgtt
3121 cctatttaaa tggcgttttt tgaatatcggg acatgttttg tcaaaccacc cataagaatc
3181 atattctgac ttttatcggg agaatatcca actatagctt ccattgaatg aataatggat
3241 ccaaatctta aacacaacaa agctttcgaa taagcatgag taatcaattg aaataaagca
3301 gatcgataag accccatacc tagagctaac atcatataac cgagttgaga cattgtgaat
3361 aggctaacc cgcttatgtc tttttgaaca agagctaaag tggctcctaa gagtactgtt
3421 attatacata tcaaagatat tatatacatt atagaacgga taacaataaa aagagggaga
3481 agatgagcta caacaaaaat acccgctgtt attagcagca tgtataagag acgaaatggg
3541 attaggaccc tccatggcat caggtaacca tacaagaaga ggaaattgtg cggatttagg
3601 aataggacct acaaataaga gaaatgcaga caaagtaaag aataagagat ttactcgtag
3661 acaaagtaaa gaataagaga tttactctat tttttaaatt taaattattg aatatttcaa
3721 acaaatcctg agattcaaaa ctgcccatta tccaataaag acctaaaatt tctaataata
3781 aacaaaaatc tcctacacga ttagttacca aagcttttcg acaagcattc actgcaatag
3841 gtctgtgtga ccaaaaacct attaatat acacaacatt ccaactaatt cccaagaaa
3901 ataaacttgg aacaaattag aactagtacc taacctaaac ttggagtatt aaaaaaccc
3961 atataagcaa aaaaatatta gatatccttg atcatgcgac atataatggg cactataaat
4021 ttgtaacacc ccaaaattca ataaaatttt gagagtatta ctaaaaataa tataataaaa
4081 atattaacat ttcataatgt ataacgttta aaaactcatt aaaactccaa atctcgaaag
4141 caaaataatt taaatgtaaa aaatctaagg ggatttgcaa aagtactatt attttactac
4201 attttgcaaa aatacacttt tctgttatcc atttttaaaa atacccttc tttgaacaa
4261 aatgacaaa ttaccccat ctaattgaaa catctaatta gaaaataaaa tttctcgcca
4321 gattttgact ctcttgtttt tttgttgtt tttttagaga caaaatgttg agattgtaac
4381 tagtatatga ttgtaaaatc gattaatttt ggtcaacatt ctaggtaagt caaacggat
4441 cagatattat tggtttactg atcagttcaa ttagaggcgt agttttaaga acaaatccga
4501 aagcatatgt agttttattg agattaaagt tttttgggtg ttttttcaga ggcacaaagt
4561 ttgggtatga atgttgagat tgtatctagg ataggattga aaattcagtc aacctatttt
4621 ttttctctga caaaaagttt ttttccagc aaaaaatgca tttaaaattc gtgtaacttt
4681 tttagaacc ctttcaaac tcgtagaaaa acctgtgaaa tttttttaa agctatgtaa
4741 agttgtttaa aaccttata aatcatttta aaaccttta aaatctttat aaagtcctta
4801 aatatatgtg taaataatat agatccccct caaaaacctt tgtacatcat tttaacacct
4861 tttacaatgg ttttaaaaga tttacaatag tttttaaaga gttttaaaag aattaacaag
4921 gattttttaa accctttag atccttttaa aatatttata aaactcttac aaagtcttta
4981 aatcattttt ttactcttaa aaaagtttgt aaatatttta aaacatttgt aactcttttt
5041 gaaccatttt aaaattcttg tgaagcctta caaaagggtt taaaaagatt tacaagagtt
5101 tttaaacagt tttaaaacc ttgtaaatcc tttcaaacc ttgtaaacca ttttaaaata
5161 ttttaaatat ccttgtaaac ctttttttaa tttttgaacc attttaaaat ctcttgtaa
5221 cccttacaaa aggttttata aagactaggt gatcacccgc gctaagtcgc gattttttcg
5281 gatttttttg taatgtttta tttaaatata atttttatt atctaatac tcaactatgc
5341 aataagcaat tactttcagt attaaaatat ttttttagc aataagcaaa tacttttagc
5401 atcacacaca taattcatca acaacgtttg tottaaacca atttaactga gggtagtgtt
5461 ggatttttag aacctaaaga tttgatgat gagctccttg cgtattcttc acctcagtg
5521 atttattata taaatattaa tttaaagaga taacttactc aaatcattaa gtaatccacg
5581 actattatac actcatactt tgagtttgaa ttacaaaaaa aacatttcga gtatgagaaa
5641 actatcaaaa gggttttctt gaaaaatcat gtaatatata aataaattta gtaatagtac
```

5701	aattaaatat	gtaaaattaa	ataataaata	tgtcattaac	ctataaatta	gaagttgaag
5761	agcttacatt	aattatttaa	taaacacttg	gcaaagtata	aagtgatgat	gtgtcaatca
5821	tatggagaaa	gttggccaac	tttcatatat	atgattttaca	agagttttta	aagggtttta
5881	aaatccttgt	aaatccttgt	aaacctttta	aaaatctttt	aaaacacctt	ataaatctat
5941	ttccctccaa	aatatttttt	ttggcggaag	aatgtatttt	cgaaaaaaaa	ctttggaggg
6001	aaatatTTTT	tctaaaattt	ttggcaggaa	aattattttc	gaatatttta	ggtgggaaga
6061	atcaatatTT	attggttgat	ccaaaccata	tactaacaga	atattttacat	aatatactcg
6121	attgggatgt	tgacccttat	caaggtgata	aggaactacg	gtccacaaca	ttaaccaacc
6181	ttaataccac	taacagacta	acataatgct	aaatacatct	attggcaatg	ctagccacca
6241	tagattgctc	aatagagcta	gctaaatcag	agtcgcacgc	tgataccagt	cttaacagaa
6301	taacaaatTT	catcattaga	atcatattca	gaataatcac	actaaaacaa	ataattcaaa
6361	atttctttct	attttctcta	tatatataat	tttgcagcac	ttttttgaga	ataatgtttg
6421	agatttgagc	catcatttaa	caattaattt	aaatggatgg	atataacccg	gtttaactct
6481	gtttggatct	tgataacat	ctatattatt	aattgggaag	cattatgtga	actacaacat
6541	tgtaagggtta	gtttaaaact	aactaaatga	tttttttgca	tggttaattt	agtaaattaa
6601	caactactcc	aactaaataa	atatcttgat	ttttgctttt	tttaaaaaacc	gaattaaatc
6661	aataaatttaa	tatctacaga	ataattgcgg	aaatcattta	aaaacggttc	cataacatat
6721	caggctattt	taatttgaat	tatatatacg	taaaacttatt	tcttttggtta	atatgtttat
6781	agataattcg	gaaacttcaa	tctaatttaa	taacaaattc	tcatttagtt	taaattttat
6841	acaattcctt	aacaaaacta	tgtaaatgtt	ttctgattag	tacacataaa	atttaactta
6901	tgatttagca	tattcaaaaag	agaatttttg	aatttttagat	ttgtaattat	ctagaaaatg
6961	gtttcagtat	atattttaaca	ttttaattag	aattttaaatg	atataaacgt	ttaaagtatt
7021	ttttttaaaa	caggtaattt	tcattatttg	aaaaatatat	acttaagcta	aatttgaatt
7081	atataaaaga	aaatatattt	ttaaacgtat	tataatattt	acaagctttt	gaaatattca
7141	ttcttctaga	ttcgattatt	tctttttttt	ttaaacagta	aaagaaagtt	taaatgccga
7201	tgtttaaaat	ttttaatgca	aatattcaaa	atgagaacaa	tcagtaatat	gattttcttg
7261	acagatcatt	aattgtattc	gaacaataat	gactaccaga	tttgcgataa	tgcataagga
7321	taagtttaat	tgaaaaacat	accaacatcg	ttgagatcta	actttagctt	tttaaaaaaa
7381	tattttttgct	atcctaaatt	taaaatatgt	aatccagatt	agtcaatccc	tgattttttt
7441	aatatatttac	aataagaata	ttttaaaata	ataaaaataa	tatttaattt	tacggaatga
7501	agttatctgg	tggaacgagt	ttgggttgat	actcatacgg	gatgataatc	tacgagtgtg
7561	atcttagaat	taacaatttt	aaaacaatta	tacaatctta	aacactaaac	aaaagaaaca
7621	atatttttaca	aaacaactat	aacttaaaat	tcgaattttt	agttacaaca	aaattgaaca
7681	acggtgtacc	gcagagttaa	gtttacatta	tatcatctaa	ttgatcaaat	tgcaaacaaa
7741	aacaataaat	ttgtatatat	atcaaaactaa	caaaaatata	caaattttct	aaatatctct
7801	tttttaatttc	gcaaatcgaa	acaaatctaa	cattcgattt	actatgcaaa	acacaaatga
7861	aatataaata	atataaatta	ataaaaaaat	agttatagta	tcataaataa	cactatacac
7921	ttttctatgc	acattataat	ttatacacta	cagtatcttt	cactaaaaat	gtattccaat
7981	acattttttta	aaatatgaaa	cctcgcatgt	atgtgcgggt	cagaatcata	tctatactat
8041	taacaaaaat	tctttctaat	gcctttttca	cactacccct	tttcaattct	accattttct
8101	attttaccatt	tttattttata	ccctatttaa	atttttaattg	accattttac	ccctattttg
8161	aaatttgattt	aggttaacaa	aatgaaatat	taatatattt	tgcttaaat	atttcaaaat
8221	aaaattttccc	gccccaaatt	tcgaaaaaaa	aattttccagc	caaatttttt	tttcaaaaaa
8281	attttttcccg	ccaaaatatt	tcccgccaaa	aaaaaattta	gaaaacactt	ttcccgtaaa
8341	ttttcttttt	ggcggaact	agattttaca	ggaattttta	aagggtttta	aaatatttta
8401	aaagattttac	aagagatttt	aaaagggttt	aaaagattta	caagagaatt	taaaagggtt
8461	ttaaaagatt	tacaagggat	ttgagagggc	tttaaaaagac	ttacaagagt	tttttaagga
8521	tttttaagg	gttttaaaagt	gttttaaaag	atttaocaaag	aatttttaag	ggtttttaaa
8581	gatgtacaag	agttttttaa	agatttttaa	agaaattata	agaaaaccct	ttaaaaactc
8641	ttgtaaatct	ttacaaaatt	gttcaaaaag	agttgtaaat	gttttaaaata	cttacaattt
8701	tttttctaa	agtttaaaaa	tgcttctgaa	gactttataa	gagtttttaa	acgattttaca
8761	aagggttttaa	atctttgtaa	atccttttaa	aactctttta	aaactcttgt	aaatcatgtt
8821	ttacoccttt	ataagctttt	aacatccttg	taaatcattc	tttacaactt	acttatattt
8881	taaaaacgat	tgtaaacocat	tcaaaccttt	ttttaaatcc	cttgtaaatc	attattttgac
8941	acttaacttat	gctttttaa	tcccttgtaa	atcattcaat	accacttact	tatgctttta
9001	aacgatttgta	aaccatacaa	acacttttaa	aatcccttgt	aaatcattct	atgtcactta
9061	cttatgcttt	aaaaacgatt	gtaaatcatg	ttttaccctt	ttataagctt	ttaaaatcct
9121	ttgtaaatca	ttctaatacc	acttaacttat	acttaaaaaa	acattgttaa	ccatgctaac
9181	cottttttaag	aatccttgta	aatcattcta	tgccacttac	ttatgcttta	aaaacgattg
9241	taaatcatgt	tttacccttt	tataagcttt	taaaaatcc	tgtaaatcat	tcaatgccac
9301	ttacttatgc	taaaaaacga	ttgtaaacca	tgctaaccct	ttttaaatcc	tttgtaaatc
9361	atttttatgcc	acttacttat	gcttttaaac	gattgtaaat	catgttttac	cctttttataa
9421	gctttttaa	tcccttataa	atcattcaat	gccacttact	tatgctcaaa	aaacgattgt
9481	aaatcttgta	aaactattat	aaaaagctct	aaaatgattt	acaaagggtt	aaaaagggtt
9541	tttaaaagtat	ttacctatgt	ttttaaagac	tttagaaaga	ttttaagaag	ttttaaaatg

1/24/03 3:20 PM

```
13501 agtctaacca aatagttatg tgaatatatg aaagagcttg actttattca ttttttgatt
13561 aaaatttagtg ctgaggttac tcgtattttg ctgattacta caaaaccatc cgaaatttct
13621 cttatagtat caaagttcag gtatttttca atcttctgga aaatatatat ttggttatgg
13681 ttttggtatt gtaatagttt atgagttttc caatccgttc atcaagattg acaatacaca
13741 tgacttctta gattttagtt tgatttagtt ggttattacc cttagcacta atttcgaaat
13801 ttatatgttg caaagctcgt agtttggttt cctctttgta ctgctttatc tcgtttagag
13861 tagctaagaa agagatagac tttagccagt taattgtctt ataacattgc tagttaattg
13921 tttttcacga tttatgtgca gaaataagct gtgatacggc acctttaa atgcgaacaaac
13981 acattttcaat tgactcggga acatgcatgt ttcttatttg tttctaccct tataggagag
14041 agctaaacca atgaattgat tcacatattt ttaatttagc agctgtcatt aatacaaaaca
14101 atcgtataaa gtcatgtctt atttgcacct ctttttaatg attccatata ggaatggtat
14161 aaggcaattg acatgatgaa ttggtttata tattatatca taggtacctg agatttggtg
14221 ttcataggtg ttcatacatc tcagtgccgt ttaagtgtac aaatcgtcga caaccaaacc
14281 aatactcata tatataaggt tagttacaca tcagctttca atcaagttta tacaccatgt
14341 ttttcgttgg tagatttggt tgactgaagg tataattgaa taaagctttc ttgtttttag
14401 cgttaatgga aatagatgat ttatgtctgc tagaaatgaa atctgtttta cgattgtcta
14461 gaaataaaaa cattgttggt tgaaacaaca atacaaaaca tcatatgaat aaccaaaaga
14521 atgaaacgaa ttccctaacc cttaggttct gaaagatctc tttgaaaaca acattcattg
14581 tttgactctg cgtgttttct tctttatcca caatgagcac tttcaatcct tttttggaag
14641 tgactcttga taaggcaaca tatagctggc catgagaaaa tactggccta ggaagataaa
14701 tttcaacttc tttgagtgat tgctccttggc ttttctttat tctcatagca aaagccacag
14761 ctattggaaa ctgccttctc ctcatccgga aaggcaattt tgtatctgaa ggtgatctct
14821 gtgcttttat gattatcact ttgtctctga ctctatctcc agtaataact ctagcttcta
14881 gtatatgatt acccatctgc gttacctgca atctagtacc attacatagc ccaccttttg
14941 gatcaatatt ctttagtaac ataaccggtg taccatctt cagagtgaga ttgtgatttg
15001 acaaacccaga aagctttata ctattgagga aataaggtgt gaagacggga ttgttcaatg
15061 acgctggatc ttgaggatca atactatccg cacttaaata agtaattgat tcacgtgttt
15121 gaaaacaatt gttagaattc ttacaacaaa atagtcttag tccacggcaa ttttaaactt
15181 agcttaccat taagtttctc gagcatgatg acaatatagc tctctcttgg aaaaatttag
15241 gtcattattt aggtcgcgac aaatcgaggt atccaaaatg ctttggtaat agcttcaata
15301 ggatcattgg catctttaat aagtaactct tcaggaatat cgatgagaac ttcaccatca
15361 ttttggtctc gaaactcttc catcaccac acgcaaggagc cagttagaga attcttatat
15421 tttctctccc ttatctttat ccaaactcatt agacataaga ctcatattct ttgtcaaagt
15481 gaggacattg caatgttccc aaagataaga tgaattttaa gcagctaaaa cgatctctgc
15541 cctaccagca catgtatca catgaagtac ttgtctgaaa tcccctccga aaacaacac
15601 cttcccacca aatggcttcc catcaacatt tccaataaca tcattttagc ttctatccaa
15661 tgattcaaaa cagtaacgac tcatcatagg agcttcatcc catatgatga gagatgcttc
15721 ctttagcata ttagcaaggt ctgaattagg tactatatga caggttgtga aatcatcagg
15781 attaataggt atacaaaacc tagaatgagc agttctccca cettgtaaca acaaaaaagc
15841 aattccgcta gaagcgacat taacagaaat taaacctttc attctaactg cagcagatag
15901 actttttcac atgaaagttt tacgggttcc tccaaaaccg taaacaaaaa aaattccacc
15961 actgttttct aaaaacagct tcatgatctc atcataaata tttttttgct cagaagtcac
16021 cttatttaaac cagtcacatc gatcttcttt cagtttttcc acgttataat ttttctcttc
16081 tgtaatgaaa cggtttgaat gatcgggtacc ttcttttagt ggtttaggca tgtgtttaa
16141 gtcttctaga gtacttccat tacacaacaa catatcttct atctctaaaa gagcataatt
16201 tattttctca gcatcactta gtactaacct tgaaaataga gaagaagata atttaattat
16261 aacagaaaaa taaagcaacg aactatacaa aataaatgta tctcaacgaa cctggattgt
16321 cggactgctt tcttttcttt gtttcaatgt cttcgattaa aattccacaa cattttccca
16381 cacatactct ggtctagcaa gacatttata cagaagcaac atagaaaata aattctgaag
16441 atgatcccca aaaaaaaaaa tttccttcga ctatagtatc aatatatgct tgatcgtctg
16501 tgagaatatc acgtgcaaaa catgcaaaaa catctttgta tgtctgtcca cgtgccaaac
16561 atgcaaaaaa ttccttcgac tatagtatca atatatcaat atttcttgt atgtctcata
16621 cacaacatct ttgtatgttc tgatatcttc atactcttt ggaccacata ctatcccaag
16681 aagtatcttc atgtaatat ctgggtccat ctttatggag gcgtagtcca ttcttctat
16741 atagaaacct ctttccctaa gcttccactg cttattctgt ccatcccatg tgaaatatgg
16801 aggaatctct gtgtaaatga actgtctagc aaactcacac tcttgattca gtttcagaaa
16861 accaataaac ataaaatctt cgtttgcttt ttttctgaa agttcatcta tattctgagt
16921 gtcttcaaaa catagtgggt gttttcctgg tagatgatat gaaagcttca ttacaggtgt
16981 agttctatc tgaatgggaa atttgaatat cctccaaact gttcagaag ctgacacata
17041 tcttttagt aagaaaaata ttagtaaaat tttagacgaa aacaaatata catagctata
17101 ttccaagtaa aaaagtatat catacctgca atcgaagtaa tctttgattt ctttatcgtc
17161 gtctatgacg gctacaaaaa gcttctgaga acctgaaaac aaaacctaa aactgattta
17221 aaccccaaat agaaactcaa acaagaaaaa ttggacatgc aatccctaaa tcgattagaa
17281 aaggcacata accttctcaa ataagaaaaa gatatgtgac tgagaagctg aggagttagc
17341 agagtcgtcg agagagattg aacaaaactc tgtagctgtg ttgagcaaga acacgaacac
```

17401	aaagagagac	agagagaaga	agaatcgatt	agagcagaac	acgaataggg	aagaacacga
17461	accgagagga	agagaaagag	tcgaatagtg	ttttttcact	tagaattttt	taaggctaatt
17521	ttttttgttt	cggcaggcaa	caaaaattaa	tccgcgctat	tacattaggt	agcgtttaatt
17581	taatggtaatt	gctatagaac	aataacccta	aatccaaaaa	catgacccgc	gtcattttttt
17641	atttcccagag	tacgagatta	tagcgttttg	aaattaccaa	acgctataaa	agtttagctg
17701	ggttataaat	tatagcagaa	catggttaac	tgctatatta	tagtgctata	aatactcaat
17761	attctttagt	tgtctttact	tttttctttg	ctgcgagttt	ctgagccagt	ctacatatta
17821	gaagtcttgt	ccttaggttc	aacaactata	gctacacgat	cttgctcttt	attaatgtac
17881	ttaaataagt	atttgataga	accggattgc	ttacaccact	caacgtttat	gtgagcccta
17941	taacgcaatg	ataacatacg	attataagga	acaacataag	tgttgtcaca	cttaataccc
18001	cctttttcaa	caaaatgttc	agattctcgc	cttctatata	ttggataccc	gtctttacca
18061	actggtgttt	gctcaaccag	tttcttagga	aaaaacttgg	aacacttacc	atcaacaata
18121	catggtgact	ccttggttggc	agcaccacat	ggaccatgta	tcatacaatc	attcacaact
18181	tcatataact	taggggtcctt	agctttatct	ggaatctcag	cagataggat	gttggtcaata
18241	tcatcagatg	ttgggaatttt	gcaactttta	tccatgaata	agagtatatg	agcatgtggg
18301	agtcctctct	tctgaaactc	cactgtgtac	atagctgcat	attaaaaacg	attttaaact
18361	ctcttcattc	tagtaacagg	gaatcaaata	aacagaatct	aaacttatga	agaagatgat
18421	agaatttaatt	accattacag	ttttgcctag	aagatgggtct	tctgttaaat	ctttcataag
18481	agagtcaagt	ttcatcttga	atatccaaca	cacaacatct	ggtctatctt	cagcattaag
18541	tttacgctct	ttacagtaac	tagttatctc	tggcaacttg	gggttacatg	taaagtgaat
18601	aaacagggttt	gggaaaccat	acatttttaca	tatagccata	gcatcatagt	aagagttcaa
18661	catatggcga	ggtccaccag	taaaagattg	tgggaattcgg	attgatttcc	cttggttcaat
18721	catggttgtg	tttccagctt	ctctcgcagc	cttaacggta	tcataattct	tacatctgag
18781	attagactgg	ttaaacttaa	tataggccag	cctattgctt	tctattgttg	tgatccatc
18841	taccaagaat	tgttggaaca	aacgcttcga	acgtagtagg	gtctaagctt	catcagttct
18901	ctcttgtatt	ctaaaagcat	accattgcct	catacttata	cacttgtttg	cttggtttcc
18961	tgttctacct	gtatatccct	tctgtattcc	tgggtctaaa	tcattctcac	cataattaaa
19021	aataagagga	tattgcattg	ggacataaga	aggatgaact	ttgcttattc	ttgatagttg
19081	tctctgtttt	ttatcttgaa	cgataatata	acgtgtatgc	attttaggac	gaaagtctcc
19141	aggaattaaa	gctgcaactt	cagacgttct	agggttacia	taggttctgc	ctcttttcc
19201	tacctatca	ctaacaattt	gcataaaaa	tgggtcccca	ttattactat	ccaacttata
19261	ttttgctagc	ctgaaatttt	caatgtatgg	attgcaagca	tccaacattc	tgatgagctt
19321	ttcaatgaga	tcctttctca	cttgattttt	gaatgttgc	ttcttcggat	tctttccctt
19381	gctaaacaca	gaattgttaa	gaaatcatat	tagtgaaaat	accataaact	tcaatacaca
19441	acgaacaata	aaatgtaaga	atagaaaaa	taaaaagtac	ctcatcacag	tagatctgtt
19501	ttccacttca	ttccagtat	caacaatata	aagctgtgag	aatttagcat	aatctcttcc
19561	accaggtaac	atactaccaa	tcaaatgata	attggctcca	tgtagctgaa	actgggttagg
19621	accttttccc	tgtggattag	atctatcaac	tttaccacca	aggggaagtc	tactgaatag
19681	catgttgtat	gctcggatat	tctcacggaa	atgccttctt	agtgcacat	cacaacttaa
19741	caattctctt	ataagctcag	gtgattctgt	aagaaatggc	aatttaactg	atccctgaag
19801	acaacataga	gagaattttg	gtttcttgc	tttcttctt	ttgttgactc	gctcactata
19861	ccaactcata	gcaccacagt	attgacattt	ccatttttga	tcacogtgat	ctaaataagg
19921	aatgtcaatg	agaggatctg	caaaaaaaaa	aaaacaaatc	aaattcgatt	ttaaaaatgt
19981	tattcatatg	ttcgttctaa	aactaaaaat	aattaaacga	tcatacaaac	ctgttttctt
20041	tttctgtgtg	ataggtgctc	ctctagtata	tctacttcgc	ttaggactta	tatctttctt
20101	acctttaata	cacgggtccc	tttgattctg	tggatgcatt	gtttcatgtg	tttcaacaga
20161	ggttttcagt	tcatatccag	aactgctatc	agaagccaaa	tcttcttgac	tactaacatc
20221	atattctctg	tcogagtcag	aaatatcttc	ttctgtttga	caacaatatg	ttattagtca
20281	aaatcgaatt	catatttcaa	agaacaaatc	tataaatcac	gattattacc	atctaaaggg
20341	attgactcaa	acagattttac	acgacagtca	tccaaatctt	tagtcttttt	agtttttctt
20401	tgtttatcog	gctgttgtga	atttttgact	gttgaacgtt	caggaaaatg	aataacttag
20461	gtctgcaatc	atatgaatat	tgatagaata	caatcattat	tattttcgtg	atggtggata
20521	attgtatact	cacggttttg	tgtgacattt	tctttgtttt	gtgtagaata	ctgaattctg
20581	cttttaagt	gtcacctcgc	tcgtcttcta	tcagggttgg	ttgatatgct	tcotagaacc
20641	cttttcttag	ttctatttag	atttgaattc	aatttcagtg	tatcaagttt	agaaacctga
20701	ttaatggaat	ctaacaaaaa	aagaacatag	atttagagtg	taattacata	ttagaattaa
20761	gatgaataaa	agctcaaagt	cagaaactat	agcaaacaga	aactcacgtt	tagtagattc
20821	tcttcttatg	tcggataaga	aggttttgtg	accaagttca	gaataagatt	ggcttattct
20881	gacagtcttt	atatttggat	gattttctgt	tgcactgtat	gctgaaaaca	gaatataagt
20941	aattatgtta	gaatgaaaaa	caatagtatc	taacaaaaac	tcataataca	aaattgggtg
21001	tagacaaata	atgctcaaat	ttgattgttt	ctgatctgat	tgaagttaga	aaatcggtgc
21061	tgccaagtat	tgaagttgac	tgttttaatt	gtatatcaga	taatatgtat	ccatcagttg
21121	ttaaacattc	aagaggttct	acatacattt	ataaggatcg	ttttgagcat	aagttatgtt
21181	ttaaaccaata	aaacattgta	aaaatagtat	gtcgatcaac	ataccagaag	aatttatata
21241	ttcctggcat	tgagagacat	tagatggcgt	aattggcaca	gagcataaag	aatatggttc


```
21301 ttgagaagac aatgattgtc taacacaaac aggtgttaga tatgagttct catcaattga
21361 gcctgaaaaa attaagagaa gccacaaaca acattgatga aaatatcgaa agtcaatatt
21421 atatagaact atatttttagc ctcttcgaga gagaatgatt tacctatatc agtcaactgat
21481 agatttttag cttaaaggaga aagagaccct tctaaaacca cagaagtaga ttttgttatg
21541 tttgatattg accgcgggca taaagagtat tgttcttgcg aagacaatga ttgtctaaca
21601 caaacagggt ttagataggt gttctcatca actgaacatt tggaaaaaaa acatataagc
21661 cattttgtac atataacaaa gtagtgaagt cttaactaca tggataaatc aattagcata
21721 ctgtagaata taatgatata catgtataag attgtgctgg aatttttcat gtaggacaaa
21781 gagactcttc taaaaccata ggcttggatt ttgttatgtc cgtaattaac cgagcaataa
21841 ttgatttcag tggataaaca aacccatttc cagatgattt agctttgtgg tgactataga
21901 aactgttttt gctcgatgaa cacttctctt tagcaaagg agtgggtggg ttttcacctt
21961 tatcaaccaa ggcacttctt aacttcatata agctattgtt tgtataacga tagaactata
22021 atttttctaac taatagagtt taaaataaaa cttagcgatg acttgaagat tatataatct
22081 tttgttaaga tatctaatta atatgattgc aattatatat tattgttata ttgattgcca
22141 aatcagaata ttattgttaa gtatctatta tattctctgc atctaataat gatgtgtgtg
22201 tttttttaat ttcagggttaa tgcttagtct acaatctaata caagattgca atctgacgac
22261 tgaggaacga ttataaagggt tagttctctca tcttcatcca atctttatta ttgtctgtag
22321 agaattctaga gctattaaca acaactctat accgtatgca aatttagatt gtaacgcgat
22381 tttctataaaa ctttgccaat aacatcaaac tttgttaact caaagggtca tgctgccaat
22441 tactaaaatt gcaacagaga caattgatct ctctcttgag agagattggt gttttagaac
22501 agaacgacac atcttcattt gtttagataa tgaaagggtg taagagttta gtattcaatg
22561 attttgagga actaagaaga ttaaaaacca aatggaaaac caattatcat ttaaaaaaca
22621 gaacaaaatg tggaaacagt aaagaaaaac cacttattaa aactaaaaca acaaaagcaga
22681 agcaaagtga acatgaatac tacttttatg ttactccat cggcttcaca ttaactccat
22741 taacttcac tgcgtactcc ctcttctgct tccattcag aaattgagct tctcaccaac
22801 gacttttagct tcattattct tttcaacct ctctgtttt atcttcacaa tcttgattga
22861 gtcaaggcac agcttcttag ttgttgaaaa attatcagca ggtgaaatgg aagtgtcact
22921 actcctcttt gaagatggag ttgtagcctc agaaggctct tcagagtagg ttagcatgag
22981 gctacactga gtctataagt agagtgatat tagtgagttt tcttaccat tagtggcttt
23041 atgaagaaac aaacaattta aaataatcca tactgatca attgaatcat gcatctggtg
23101 tctcttcttc cccaagacca tcataagcaa gaattctaga taccttataa gtagcaggtt
23161 cattccacac atttctattc tcaacacaga ccagatactg ataagtttta ccgataagat
23221 tcttgattga atctggaagg atgtctgggt cctcaatctc atcgaaagat ccatcaagga
23281 ggttattttg tggacaacca acaatctcag aagacattgc atcaaaaagc atcagcttca
23341 tctcaccact tgaatccatc actttaacat gtagcttgta cctgtatagg tcacaatatt
23401 taatccaata agccacagac aaaaatcttt ctatttttca ttaaatcaag ttatgagac
23461 ataccttgca agcacattgg tgattgtgct tcggcaagtt tcacaccata actgtgcttt
23521 caaggaagtc atgactacct tagtcacttt ctatttgac ttcttgacga caatatagta
23581 ccaagaccac tcctgtcaa tgtcataaat tgtacagaga acctgaccc tacciaacctg
23641 agattagaca attataagca cgattaaaaa ctatgataag tctaaaacca cgtcaataga
23701 taaagaaaac ggatattaca tcaaacatct cagagacttc tttgatggct tttttagcaa
23761 actgaagata gaaatcccca gtaggaaact catcaatttt cgactttggt tttagactcaa
23821 tgagagttag agcaagacaa tcttttggtg aagcgtcttt aaattccaga atctcaggga
23881 gatcaggatt gatcaagatc tgagatgtat tgaaagcatt ggagatactc ctaacatcta
23941 caaaacaaaa caaaaggatc catttcatta aatccgtaa caaatatata tatatatata
24001 tatatatata tatatatata taaactttca cgatatagtc aatttagcca acatatacat
24061 tctatataac ttcctatata acctttgtaa gtcttgagct tggcaaatct gaccaaacat
24121 atgattcatt caccattagc agcttcacag gcagtgaata cttgttcagc aatgaaccoc
24181 cagagagtac atggtaacct tctatcccta aattaataat agaattatta gaaaaaatc
24241 aaacttaata ttttgatttc tattactcat atccaagtaa acttactttt gatccctcag
24301 ttcgaaatca atcttctttg ttggttgatt gtggacatca atagtctcca tttcaccaac
24361 attcactatc tgaccaacca catctaatac aaaaacgaat cacagattca aaaaaacgta
24421 atggctataa agtttcttat ttaaatggaa acgaaataag aagcaaaaaa cataccaatc
24481 aggtagtttt ggttgagacc accattttaa acatcattga aagcagtaaa ggcagaaaaa
24541 taagaatctg acaatggatc gattctcaca acacgggttt gagccatgaa toccattttg
24601 taacgatgat cagttgcttt gaattgacct gttgcatagg tcaatccaaa gttttcaata
24661 gccctccact ctcttggaac aatgagcctc tgaacctgt tcaaatgctc tctcttaact
24721 gaggcattga tcttctttcc cttacattaa aaagcataaa atattagtat agaataaacg
24781 tttttatttg ggaatattgt caatttatgt agaaataaca tgaatactta ctctcacatc
24841 agagagaacc atgtcgaaag tctcaccaaa cttcattgtt aacgttttcc aggtatgaag
24901 aacacagact taaactctcc aagacgtttt gtagggcttg atgtcatcaa ggttaagcaat
24961 tgcttgagcc attgttttga tttagttgct agaaaatgct gtatttgagt gatgaattgt
25021 gatgagaatg gctttgatat atatagtggg gaaggaggct aaggttgctg gatccaatat
25081 ttcaaaaata ttcgtcagta tcttgagtag taagttgcaa atcaaagatc ggataatgac
25141 taatgatctg aaattgggtt agttcataat aattggccat ctatagataa gtggttaata
```

```
25201 ttgttttagtt gattaaggtt atttctagat ttaacagtat aagtttatat gattttctag
25261 tttatggaat gtattatcgt caatgaggag ttagaataaaa cgaatatgct cgtttatgtg
25321 gattttatact gattattgag gagaatgaat atatgattga agattataac aatcgttaatt
25381 gaatatgcag aagatTTTTga aggaatatgc gtgtatcact tccacaaata ttgttgagat
25441 atagaaataa acgtgttccg tgaaagattt tgactaagat tgcatttttg atgggccacc
25501 gaatatatag gttttgttgt gatcaaaagc ccattaaaca taacacaccc tacttaataa
25561 aaagcggatc ctttttactt atgttttttg cttggacaat tttacccttt agcatgaaga
25621 acaaaatcca aaattgaaat caaaattgag gacaaattcg acaaaattc aatttgtact
25681 tctcttttag taataagagg atacaatcct aaacactaaa caaaattaac aaaacaaata
25741 cacttaaaat ttttgatttt ttaaaacaca tcttgattct tttataagaa gtttaaatat
25801 tataaatatt taaactttat attatgggtg aaatttttag atagactggg ttgggttgat
25861 aataatatgc gatagtatac aaggagtga atcctagaat gacaataaaa atataactat
25921 caaatgaaat actaaacaaa aaaaacaat agtaacaaaa taaatacatc gtaaaacttt
25981 tttgttaaaa cgcactctga ttctgttata aaaaagttaa acattataaa tatttaaact
26041 ttatattaca ggtgaaatca tagaattgat aggtttgagt cgattttaat atgggattgg
26101 gtactacggg tgaaatatga gaaaaataa tcaaaataca attatacaat attaaatact
26161 taacaaaata aacaaataca ctttaaaact ttaatctttt agttataaaa ttttgtctca
26221 cgggtgatat tacattttaa tcaacaataa gcacaaattt atagatcatc ttaaaaaaaa
26281 ctctattatt ttttatttaa tgaaaatata ggcccgcggt ataccacggg ttaatatcta
26341 gttcccaaat attatacaat tatacatagt cttaacagaa taacatatct tactagattt
26401 taaccgcggg tataccgcgg gacgattttt tttttaagg taatatatat taaaacttgc
26461 aaattttatt tttataaaat atttatattt tacagtttat aattgttatt aagtaacggt
26521 atacgacgaa tcaatgagac aattgtttaa aaactgaagt tttaatcctt attaaacaga
26581 atattaataa tattttaaaa ttttataaat attgattcaa atacattcac catataacct
26641 aattccaaaa taaaacctgt tctctaattt ctttgcccat ccataaattt tttttaaat
26701 agtaattttt aaaatttaag aattatttat tatatataaa agttttgcaa attgtatcat
26761 tctctaatac atttatattt tatagttaa ttttatatat agtaacatta tacataccac
26821 ataaaagttt ttggttttat ataacttttc taaattagaa tgatttgata tgtttaataa
26881 gtggtcttaa taaaataata aattaaaaat ttaacctgta ttgaaatggg ggattaattt
26941 attttactt ttttattaat gttgatttaa aaattcgtcc ctccaaatcc gtcttgcaa
27001 aaagccaatt attttattaa tattaatttt atttatgata tgtctctgaa ttataatctt
27061 aacattttag ctaataacat atgaatgcac catattcatt taatagggga atgcaattat
27121 aatcgaataa gttcataaaa caatttaata ttttcatatt gtttataatt tggtaacaga
27181 tagatatttg aatattcggg atattttgct tctgtttagg aaatctttga tgatccgtct
27241 ttagatacaa taagccctat aattaagaaa tgaatattgt agttaaatga ttttggtaac
27301 cttaatgacg tcattaaagg attttgtttt gatttaaatt tttctaattg atctgtcatg
27361 taatttaatt caaaaattaa ggttacattt aaaaatgtact tccaaaataa tatagtagga
27421 ttattataat tataacccaa ttaattatga aagcaagcaa ataaattttc agaataataa
27481 aataattcat ttattctttt tatcatcac ccaaaatatt atacaaatgt ataataacta
27541 ataaggcagt taaaactcac ctgaaaaagc ttttaaaacc gaagacaatt ctatccacgc
27601 tccaaatcac caattaattc gtgtaaattc aaccgtgaat caaaaacata tgtactatat
27661 accatccaaa agctaatttc aagttcaatt gtattctagt atcgcgtaaa ccagtttaaa
27721 actagttgaa ccggtcttaa ttactaaca cttaaacaca aaactcaacc aattccaaat
27781 tagatttcaa acaagtattc tccatagaca gaacattcaa gaatcggttt taaaacctta
27841 agaagcagcg taaataaatc ccgatttgac tatgaaccaa caaaaagaaa tatttaccga
27901 tcgaattccg ataaactctg acttccgatt gataatatgt cagtggcggt ggattcttac
27961 gaacgagaga tttctaacaa ggtatagatc aactccta at tctttgtgaa aagagatcaa
28021 caattaaaga gctttaaagg tcattccatg gaaaattttt gaaaaagggt gaagatgaa
28081 gcttacctag aactctgaag aaggctgaaa acaactcaat actaaagaac caacaataga
28141 agccataacc aacctaaaca accacaaaat ccgataacaa aatatcaaca aatattcaaa
28201 tccgcattga ggcaaaatgg gtctaagaaa ctccacaaaa gtctcaaatt gataatggat
28261 toggctttta atttcgggtg agctttaaaa cgcttatata tacgatatat tagggtttaag
28321 gaacaaggat tcgacttgc aaaaaacttcc cttcagtttc gggaaagata aactggaatt
28381 ctcttctctt ccacgttttt gcacttttagt cctcaaat tctctattt acagttccg
28441 cctaaatttt tgaattgctc agaaatcctt gtgaaatcag acagtttcaa tacgttcaat
28501 ttaatccaaa atctgatcct atattttctt ttaatttttg gtgttataat ctacgtggaa
28561 gaaattgtga gattgcatgg tgtcgtgagt accattgtct cacatcatga tccaaacttt
28621 acttcacatt tttggagtag cttacaccac gtattggata cgaaactaca ttttagtatt
28681 gtttatcatc cttagacaga tggacagtct gaaacgacta tacaacact agaggatct
28741 ttgagattat gtgttttga tagtcatcag tcataggaga ggtcattgtc attgattgag
28801 tttgcttata ataatagctt ccatgcagat atcgtaagc accatacag gcttgatagc
28861 gtaggccatg tcgtacacct ttatgttggg cagaagctgg agaaactact ggtattggct
28921 ctgagttaat cttaggaacc actgaaaaat tgaggctgat cagagatagg atgaaaccgt
28981 cgaaggcgtg atttagaatt tcaggtagac gatcaagtat ctattaaagt tcaacctatg
29041 aaaggcaaa ttcgatttgg acagtcagga aagttaaaac cgcgttttat aggtcctttt
```


1/24/03 3:20 PM

```
33001 atatatctta tatgtatatt taaaacaagg agtagaagtt aaacattatg attattcaat
33061 aatttaacta attaataact gtgaacttgt ttattttatat ttttaaaaaat tagctgttac
33121 taaccaatca caaagttaat atgtcaattt tcacaattta acaagattca gccatataag
33181 ttttaatatc aacaactaaa taacatgagt tttgttttcta tgataattaa tttgggtttt
33241 ttttttgatg aactaaaagg aattttggat taaatcttga tcgcaatttg ccaatcatca
33301 taattcttcc catcaaatat ttcagaagca atggaaactc cgggattatc actgtttgta
33361 aggtagtaag gtgaatatat gtataatact gtcattgagt tcgattactg gatttcgagt
33421 acgagctcga ggttcacgtg aaccttcagg agaagaatca tcgcgaattg ccaatttacg
33481 agctgaaaca ggaacgcgga tcgtgcgatg attagaaatt ttacgtccat gcgccattaa
33541 aaaagctctt attttacggc gggggataaa atttgttcaa ctttttggtc gatgcaacaa
33601 atttctcggg gacgaaaccg aagaaactgt tcgaagagaa gaagaagcaa agaaaaaag
33661 aagtcaagaa tcgtatgaat gattcaccat gttacaagaa aatttgttac gaagaagaag
33721 ataagaatct tatatattaa acagatgctc aaggttcttc acgttcgctc tcagaaaatg
33781 aactgacgga tttggccctc cctccactt aaggttttta aggataaatt agtcttttta
33841 ctaattaatg agataacata ttattttta atgtagaaaa taaaaaaaag cgtttaattc
33901 cctatcaacc cttaaatttc tattctccag tgccccta at caaacctaaa tattaaccct
33961 aatcaaacct actcaatcct aagtcgcccc acaaccctaa cggacttata ttatgcccg
34021 ataattttac cgtcatatg taaaacccta aaatcttatg ctcaaaattc ctataaatat
34081 aggatctctc ctctcaaa caaacatcgt agcttgattt tagtgcatag caatagcaat
34141 atgggtttata gttctttttt gtatttcaat caaggtcttt tctatgatgc tctcatagc
34201 aatagatttt ttcgagtcct ttttatgtta tgggtcaattg gaaagtaaga tctcatagtc
34261 tttgccattt ggtgatgagt tgtttaccca aagtttatag tttgttcgat tacatgtttt
34321 ggatctatcg tcgctactat atgatataat tattttggaa tttctttctt ttaatgttag
34381 tggatcgatc cgtcttttaa agtgatgggt tttgtgtatg attttgcaag ttggtgttta
34441 tgcaacgtcg taaattgggt ttatctaaact aaatctcctc aaatctgagg tttgttttaa
34501 tcctgcaaaa aggtagtctc ttatctaaact aaatctcctc ttttatcatg gatttagttt
34561 ttgctttttt tttttctggc tgctataaac cctagcgttc gaattgattc acacagatca
34621 tataaacaca tcatataaac acaaacattg tatgtaagtt attgattttg tttatagttt
34681 gttcgaattt tttgttttct attggttttg agaagaaata tgcataagaa tttgtcaatt
34741 tagctgaagg aacgtttatt tgttgttgtg cagtcttaga cattgaggcg gtaggaacat
34801 tcaacatgtg tcacgcagct ctcaagtatc ttaagaaagt agcgctgga agagactcat
34861 caagcgggtg cggttcgatt attaacataa gcgctacttt gcactacaca gcttctaggt
34921 atcaaataca tgtatctgca gccaaagtaa tttgattagt ctgataaagt cattcagtg
34981 ctcgagtttt ggttcggaaa tcaagaacta acttaaagga acaatagaat gtagcaaaat
35041 gtgatcttag gagtttgaaa cttgaaacag gttgcagttg atgctaccac aagaaacttg
35101 gcattggagt ggggaactga ctatgatatt agagtgaacg gaattgctac aggtcctatt
35161 ggaggtacac cgggaatgag taaacttgtt cctgaggaga ttgaaaacaa aaccagagag
35221 tacatgcctc tttataaact tggagagaag tgggatatcg caatggctgc actctacttg
35281 agctgtgatt ctggtatgct gagttaaata aagcacactc tttagtataa aacatatata
35341 gatgggttaga tgacgtaact catttcttga tttgtttaac agggaaatat atgagtggac
35401 taacaatggt ggtagatgga ggactgtgtc ttagcaaac ccgcatattg gccaaagagt
35461 ggagaagagg tctagggcca ggccaatgat tattactatg ttttggtgca tattgtacaa
35521 agaattgtat attaaaacca aaaaagctct tgggtgcttg gtgcataata attatacaag
35581 aaaattctca caacgtctac gaggaagcaa aaacgtctaa gcctaacc aa gatgaagttt
35641 cagaccagtc tctagctctg aatacgtaat tatggtcaaa actcaaacgt atggccgagc
35701 tagaagaaaa gtacaagtcc cttgactcaa aatcagcaga agatgttggg aatgtcctga
35761 aatcagttgt actttactct agttaaagtt gtaatgtcga ggtttgacat ttgtataata
35821 ttaggtgggt ataattgtgt gctagagacc ccttataatc tatgagtcag aactcattgc
35881 taatattata caaatctaca gggtcacaca cctaagcaaa ttggatcaat aaaaagattt
35941 attgcgcttt ggccattat taacccta attgattagggt taaacgtttg gctatatata
36001 tgatatatta acaatttttg tgagatacac ttttgatcac aagaaccata aaataaagag
36061 acaaagagtt gtctaaagtt ttgagattaa atcttaaagg aactagcatt tcagatctga
36121 tccagtttgt gtgcgtgtgg atactagtag aggcacaaca cttgggggtgc taaaataatc
36181 tcgagataat atggttcacc tttgcgctgc agaatatcca ggtatattca aacccaaaga
36241 tctaaattta attctttatt tacacagttc ctggcaattg aatttgttt tagtatttat
36301 taattgttat aaaccggtac gtaattcctt cagtggatc agagcctatg agtttaatat
36361 tgaatttgaa tctagtttaa tctatattat aatatagata aataattgtt atttgtttat
36421 tatgatgaaa tttgtgaaat ttaatttggt aaacatggta gaaacgaatt cgctaataag
36481 aacatgaccg ttgtattttt atgggcagca aacgaaacca acttcttctt tgtagactat
36541 agatttttgt ttgaaacccc tatattttag aaattttgtt gctgagaat aattatttat
36601 gtatttgtag gaggttaatt gacatatcag accactgtaa gaagagagct aacggacctt
36661 aatgtccgga aacggcaagc gtattcaggg tgtaagtagt ctctgattca tctctctctg
36721 atttgatttg gtacagtata aaattgtcta ctttttttta tttttgattg actcgtattg
36781 attgtgtgga ggttattgta ttgattgatg cgaatgtagt agtttttttt tttttttttt
36841 ggtcggaaaa cgggagaagt ctcccgtctt attattcata acaaaaagaa aaagaactac
```

```
36901 aaccgaacaa aacgtaacct tgagacccca ttggcatctt ccaacaaaat cggattaaca
36961 atatccggtg gaagattaaa agaattgaaa cctaaagggtc tttaaaaaac atagttcgct
37021 aaccatcccg ctacacgggtt agcctcccta tacacatgtg acacacggac taaccagtc
37081 cttgataaaa agccatagca caagcgctact aggaacgaca ggggggtgaga gtcactgatc
37141 cctgaagtta aaaaccccac cacaatctcc gagtctacct ctagctccaa cctcgcgacc
37201 cgcctctccc aagctatgta aagtcctatag tagacccccc acagttctgc caaaggagct
37261 gtacaaatcc caatatgaag cacaaggcct ccaatccaat taccttctct atcccgcaac
37321 gcaccacctg cagtcgcctc tccaggggttc ccatgcgacg ccccatccgt attcagcttg
37381 acccagccct catttggttt cctccatgag acttgctcgt ctacctgcac aatctccttc
37441 gtattcccggt agagctttta gtgagcagct gaaacctcat ttgccaaatc cttaatgaac
37501 cgtaatctat cccgacagcg gccaacctct cacaacacat ttccacaacg ccatctccaa
37561 ccccaccaag ttgccaatgc aaaaactgtg gaccaaacac tgccctccat ctcaccgtgt
37621 tctcccagat ttacatacag ccactcaaga atagacttct caaagaaaag gcgttggttg
37681 ctccaccgaga taagccgtag ccatattcct tccattgcag ggcaatctct gagaacatgg
37741 ataatagact ctttcgctcc cttacaaact tgacagatgg tcgtggtact caaatgtcga
37801 cgatgccttt cttgattcgt cataatggct tgttggtgcta ctaaccagag gaagagtctc
37861 actctctcag gcgcagtcac gctccatact cgggtcaaaga atctcttcat atcttgctc
37921 ggccgctcat ctctagtaag aaaagaatag gtggtgctca ctttgaatct ccatcttga
37981 ctctccccc acgatattct gtctctagct cccgttatat catctaaaac aaacgattgt
38041 aattgtaatc tgttataaga agatatatat ggagctatcg tcgccatctc ccaaccgatc
38101 ccattttctcc taagatccct agctcgcatt cctttccaac ccgccggagg ctccctgaacc
38161 aacaactcag ctagaggaat atttgttagc cacttatcca tccaaaacag aatctccga
38221 ccattctcaa tcacccaact taagcccggg attaccacct ctctcatacc caacgtaca
38281 cttctccaa gctatgagca gacacttcta actttagtcc aagcaccatc ccgacttct
38341 tgtactctgt agttacatct catgattctc gcccataagc tatgataatc ctgaattaga
38401 cgccatccta ccttagagag aagagcttta ttcatatcct gagctttcct aattcctagc
38461 ccaccttcac tcctaggctt acaaaactct ttccaggaaa taagatgctg cttcctttga
38521 gttacagagc tgcccataaa gaaagatcta gacactttgt ctaaccatc caacgttgat
38581 ttccggcaaag caatggtgct catggtgtga accggaatcg acgaaagcac tgccttggtg
38641 aattgtaact tcccgcgaag actaaggaaat ctgcccttcc aaccagccag ccttggtgtc
38701 aatttttcca agatatcacc gaaagtgtcc ttattaattc tctctgaag gaccggcatt
38761 ccaaggtatt ttccaagctc tctagttaga ctaatcccac tttcatcgt tatcaacttc
38821 cccaagtccc gagacacatt ttcagagaag aagatcttcg atttctctag gctaactttt
38881 tgaccagaag caacacagaa acgttccaga actcgtctta tcaccogaat ctgagccaca
38941 gatgcttccg caaaaaggat tagatcatcc gcgaaacaga tatgtgaaag ctttggtcca
39001 cctgagaga gactaatggg ttccaactct ttccgagcaa tagaatgctc ttaagatgg
39061 catagctact ccatgcaaag cacaanaaga tacggtgaca gtggatcacc ctgacgcaac
39121 ccccggtttg gtctaaacgc ttctgttcgc tcaccattcc acaataaact cattgatggc
39181 tcagtaacac actgcatgat ccatacaatc cacttctctg gaagcctaac tgcccgtaga
39241 gtatcctcca aaaattccca gcgaatccga tcatatgctt tttctaaatc caatttcaag
39301 agcatccaac ccttacgccc ttttttccct ctcatcgaat gaacagcttc ttgcataatc
39361 actatattat cagctgcca tctaccggaa ataaaactag cctgagccgg cctgataagc
39421 ttcatataaa cccgcttcaa ccgcacacac aaagttttag taattgtttt gaataaaaca
39481 ttacataaac taatggggcg aaactgtgtt atcctttccg gtttcagaac ttttagctatc
39541 aacaccacga ttgcatcatt cgtaccttct ggaagtctcc cggaagcgaa gaagtccaac
39601 acaaacggac taacagaatc acctactata tcccaacaat cctgataaaa aacaggttga
39661 aagccgtcag gccctggagc cttataactc cccatgcttc gaattgatgt ctcaacctcc
39721 aagggtattaa acggcttgct caaattagac atgaaaggcc cggatccgga cgcctaaag
39781 tccggacgag ttctctcgac ccaagccact aagtgactcg gccctaccgg tttatgaggt
39841 tctttcttag tctaactcgt ttcatatgat tccgtgattc aatatatgag acgaatagag
39901 taacaaaaca atcggatact ttattgatta agcaaaggat tcgatacaac ctatgagttt
39961 gcatataagc tattacaagt catataatcg atccctagac tatccgaact atcaccaccc
40021 atcatcccg ccttgagctc ctgcgccta agccaagtc ttcccacgat caccattcgt
40081 tcctgaaagt catagagcca taagtaattct cactttactta atcatatgca agtccatgca
40141 atgctctata tgccatacac tactcggatc cgtggagaga atttagagag gggaagatcg
40201 cgcacagtc cggccgattc tcggcagggtc agttaagacc tgccgggtcct ccgccgagcc
40261 gcgtcccccac ctccagcgat ccctaccggg accaggtctt gtgtaggatg cacatgaccc
40321 ggatactacc cagccctcta gtcgatcctt tcaaggatcc gactactaac tcacatccta
40381 gaacccttag actcaaagtc taggataccct taacctcctt ctcgaggtta ggtcgatccg
40441 accctccttc ttatttgccg gccacagccc ggctaccgaa tcctaagact atttattctc
40501 ccaagaggga gagagtttta atggtgagtt tcttggtgtt tgggtgtgtt gtggtgagtg
40561 gaatggctct gggcgccctt ccttatatag ggggcgaatg aatgcttccc gaatcaacac
40621 atccgttctc tgcaccttcc cgaagagttg cacttttccg aagagttgag ccctcccgaa
40681 gagttgcacc tttccgaaga gttgtatctc ttccgccata accgcccgag tagtccgtca
40741 cttgtccagc tacctcgagc tcaaccgtca ctcgccagc tacctcgagc tcaaccgtca
```

```
40801 ctcgctccagc tgcocgagct ggactgtcac togcacctgag taaccgtcac tccttccagc
40861 tgagcttgag ctttgggoga actgacccag ctgatttctt tacgatctta gcttagcctg
40921 catgatttcc cctttgtccg agcttatggc cagctccttc tcaattgtcc gagactccac
40981 ctattttctct cagacccaagt cctagccaac gatcctattg aggatcgcgg gcgttacaag
41041 tctccccctt tgaattggat tcgtcctcga atccgcacca agttgatctt taccatctt
41101 cttgaaccac tcgggaaaat tggccttcac cttattctcg gtttcccaag tgtattcttc
41161 acggcctcca caattctaca agacttttaa caaatccctt gattttcccc gagtcccttt
41221 cgacattcgg tccatgatcc gtaccggcca tgcctccacc gtcattgtct cctttaaccc
41281 gggaggtata tcttccacgc tttcttcttg atcacttaaa tacttccgta attgtgacac
41341 atggaagacg ttgtgaaaca catttagttt tggaggaagg tcaagtttgt atgccaccgc
41401 cctactcgt tcgatcacct tgtacgggcc aacatacctt gggcttaact ttttcttga
41461 ggtgaatcgc cctgcccctt tgtaggtcat cgcctttaag tacaccaaat cctcacttg
41521 gaattccaat tcttttcgac gttttctggc gtatcctttc tgtcgatctt gggcttcctt
41581 taacttgatt ttttaagaact tcattctctc cgtggtttca tccacgatcg taggtccaaa
41641 caaccgacgt tcgcccacgg gcgtccaaca caacggtgtt ctacaagctc ggccatacaa
41701 cgcctcatag ggagacatac caatactagc ttgaaaacta ttattgtatg caaactctat
41761 taaccttaga tatttctccc agtttctctc ccaatccaag acgcatgctc tcaacatgtc
41821 ctcaagggtt tgaatcgttc gttcggactg tccgtccgtc tgtgggtggt aggcagtact
41881 taggttcact cgagttccta aggccttttg gaaagcattc caaaaactttg acgtgaacct
41941 tgtgtctcgg tctgacacta tactaaccgg aatcccatgt aaccgcata tctcgtctat
42001 atacttctcg gcaatgattt ctgctccatc cttatccgag atcgccatga agtgcgcaga
42061 cttagtcaat cgggtctacca caaccatac cgcattgtgt ttggatttga tccccgttgg
42121 cagtcgagtg acgaagtcca tcgttatgtg gtcccacttc cattccgata taggtaggtt
42181 ttgtaacaac ccactttgta cttgatgctc cgccttgacc agttgacaag tcgggcactt
42241 ggccaccat cttggccacgt ctttctcat tccataccaa tggtagtaac ttttaaggtc
42301 gcggtacatc ttgttcgacc caggatggat tgagaacttc gattgatgcg cctctcttag
42361 gatctctctc tttaatgccc tatcgttcgg cacacatact cgaccgttca ccacgattgt
42421 accattgttc gaggttttgt actcggctct gttgtttagt gtccatcctt ttatctcttc
42481 gtcccgttcc tgagccagcc ggatcctact aagcaagtgc gctgatcag ttggtcccaa
42541 ccctaagggt tcaacttctt tagacaaggc attcacatgc aatgtcccca tcatgttaac
42601 caagtccact tggctcogtt ccgctcaaac ttctagcgc cctcactca ggcgtcggc
42661 gacttggttg gccttgcccg gatggtacgc tatatccaag tcgtaatccg ccacaagttc
42721 catccacctt cttagcctta gattcaactc cggctgggtg aatatgtact ttaggtttt
42781 gtggtcagta tagatttaaa ctttagcacc atacaagtat gatcgccaaa tcttcaaggc
42841 aaatactacc gcggccattt ccaaactcat cgtagggtaa ttctctcat gtttccgcaa
42901 ttgccttgat gcgtatgoga taacactccc cttttgcac agtacgcac ctagtccac
42961 ctatggatgc tcogtataaa cgtataggg ctacacttcc tccggtaaga ccaataccgg
43021 tgcattaatc aacatggcct ttaactccag gaagctttc tcgcaactc cgcaccagtt
43081 gaatgcggtg tcttttccgg tcagtcagat caagggttg gccatgctcg caaaactcat
43141 tacgaacctt cggtaataac ccgctagtc gagaaaactc ctaatctcgg tggcattcct
43201 tggctcgtgg cactccttta tcgaccgtat tttctccgga tcaaccgata ctcttggtc
43261 cgagataaca tgaccaagaa aaccaacgct cctttgcaa aatgagaatt tacttaactt
43321 cgcataatag ccatgttccc ttagcgttcc cagcaccgcc cttagatgtt cttggtgagc
43381 ttcccaagat ttcgagtga caaggatata gtcaatgaag atgataacaa actcgtctaa
43441 gaaatcccg aatacccat tcatcatttt catgaaggct gctggtgcat tggtaaatcc
43501 gaatggcatg accacaaact caaagtggcc ataccttgtc cggaaagccg ttttctcac
43561 gtccgttggc tctatgggaa tttggtgata tcttgaggcc aaatcgattt ttgaaaacca
43621 ctgtgcacct ccaagttgat ccatcaactc atctatccgg ggtaaaggat atttgtttt
43681 caccgttaact ttattcaatc cccgatagtc gatgcatagc ctaaaagctac catctttctt
43741 tttcacaac agaaccggcg caccocatgg tgagctactc ggctgatga atcctttagc
43801 caacaactcc tccaattgtt tctttagctc cgccatttcc gccgtgcca tccggtacgg
43861 agcttttag atgggggttg tcccgggctc gagttctatc gtgaacgggt ctgaccggtc
43921 cgggtgtaca ccactaacgg ccgcaaacac gtccgagaac tcgttgacga tcagaatgtc
43981 tttcaactcc gcactcgccc ccaactcctt ggttggtgatc gtggctaaa atgcctcaca
44041 accctttccg agcatccttt cagctcggat tgccgagatt accaagcttc ccgaggttgt
44101 tcttattcct tgggaatttta acataccctc gtccctctcg aattgtatcc gttctcgatg
44161 acaatctatg tgaccttgt acttcccaaa ccagtccatc cctaagataa catcatgttt
44221 ctttagtggg acaatgatta ggtccgcggg catattgaca ccgttcaact caaccgagat
44281 gcccgaaca agtctggctg gatacattgc ctgtcctcgg gccgtcgtat ccattccata
44341 atcgggtgtt ggttcttttc tgaaccaccc tgtttcaacc aattctagac taacaaaaca
44401 atgtgtagcc ccagaatcaa atagagtgtg cgttactacg tcacctactc tcacggctcc
44461 tacacattcc cttagaatgc tgagtacctc ccctttccca tatgcatgat gcaatgcaaa
44521 tgcagttatt gaaataatga ataggaggtt gtgagaacgt accagtgatc gccctgaagt
44581 ttccggcacc gttagcttcc ttgcacaact cataaactcg cgtgccacg gcttgccgt
44641 ttgggggagg cggcagttcca tttccccgcg gattgtccct ttgtccggcc tgcctctccg
```

```
44701 cagtcgaactt agggcactct cgcttttagat gtccagtcctt tccgcaatag aaacaagtcc
44761 ttgtctcttcc accgagaacc ttggactgtc ctgggttccat tcttggggcaa ctttggatcg
44821 catgggcctt gcttccacac cgccacatg cactatggc ctttcagcaa gttccgctgt
44881 ggctcttacc acagggtcag cattccccgg tcttagcatc ggacttgggtg ttatccacct
44941 tgtcaaatatt cctttttctta tccgcgggtt tgggtgggttt gttggtaacc tctcaacaac
45001 ctcatcgaca tctccatag aataaagacg tgagtagtaa tcgacagcca tcttctcaag
45061 ttctgtctgt tgtgaaaccc aactcccatc atccgctttg agcatttcaa ttctgtttct
45121 ccgcctccta acaacagtca tagtatggaa ataactttgtg tttcgggtctc ccaactctac
45181 ccactttctct cgggattttct gaaaccatag tacttccctc tgctcaagga ccacatcaaa
45241 ttctttaatc aactcctcct ccttacttaa gagattatca gtttgggttaa tttccaaaag
45301 ctcttgaacc accttaattt cattcatcaa actctccttc cttctattca cgtctccaaa
45361 gacttcaagg ttccactttt ttaatttact ctttagagct gccaatgcta caggcgtctc
45421 tcttccgcta ttccaagaag cctgaagcaa atccttaaaa ccactatggg taagccaagc
45481 cgcttcaaac cggaagggtc tctcagcgg attactcgc acctctgggt ccagctgaat
45541 ataaatagga gctgatctg aagcgaagaa cgggaggtga gtaactgaag ctctctgcca
45601 ctttaaacgt gtttgaggac ggcatagcac cctatccaaa cgctttgcga caaaagtact
45661 ttctacacga cctcgtctcc atgtgaattt gttccctttg aatcccatat ctatgagaga
45721 gagctcattt atccattccc cgaagctag tgaatccggc gagagtcctc ctatccccc
45781 gttctttctg cgagtctgat aattgtgtta aaatcacctc ctacaatgat cggctcatcc
45841 acactctgca ctatccggt caacagactc catagtccac ttcttctact aaccgaagga
45901 gccgcataaa ccgcagcag atggatagcc gctaaccctt tcccacttt cgcgtgtatg
45961 aattgttccg ctgactccac aattgaaaca tctccacct ctgatttcca caagagccaa
46021 atacctccac tctgcccctg cgcactact ctgaaagaat tttcaaaacc caaattttga
46081 taaattctga ccgctttctc accaccgca tgtgtctcaa acaaagccag aacatctatt
46141 cgatgtttct tcagcatata ccgaattgaa cgacggaagt tgggtttgtt gacccccgg
46201 caattccact ttaagcaatc cataagatat ctgtaaatca agggcaaacg aatcacggg
46261 gcgacctgtt atacaagggc agagcttccc gcttcaggcg acaagcttc attcacacct
46321 atccctccct ccactcgagt ccggctgaa atcaggctcg tcgtatccat ttcatctctc
46381 cggaccgata acagagcact ctctgtgatt ttctccatct cagtcctggt cacaaaaatt
46441 cctctcgtc tcccctggt acctttctct atctogaagc ttttccctgt ctgagccgc
46501 actacctcac ctttctgttg gccaaagaca agccttttag tgggctatt cgtgttctgc
46561 ttggcctttt gacgattctg gtcattgggc ttaaagaatc cggccctct atccttcagc
46621 ccacctttca gcccaatttt ctttatctca agtctccac catcgattcc ctctogaatc
46681 ctttctttat cttgactaat actcttccct tgaataaact tacgcagagt attaacattt
46741 tctttattct gaactccga caaatccgc tcaactccct ccactggaat tgattcgata
46801 tccaagtta caaaactatt cgataccaga atattcccg aattttttcc ttgagtgtt
46861 tcccgaagt tgcgcctac aagcccgta gggcctcta tccattgac caccgggacc
46921 tgcggcgggt cgccttctt cctcgacct ctaccaaccg taaaaccatc attctccgga
46981 actgctccac taaccaccac cggagcactc tgagaagcgg cattaacttc tttctctgca
47041 atagtctgtg ggcagggtat gatcaagtga ccatacatgc cacatcttga acaaatatta
47101 gccaaacctt catacgagac aaaatatctc tcaccattga gtaacaaagt tcccttcaac
47161 ggcttttgca aatctacttc aatacacatt ctagcaaagc gggctcgtc tacatgcagc
47221 gttgtcatgt ccactttcac cggctttccc aaagaccag ctataccat aagaatggac
47281 gtgtgataca gagacaacgg gatattcatc agacggacc aaattgggtg tgtgtgatc
47341 tcatcgcgta aagggtcaaa ttctggagac caggcttgca cgagaagata actgccaaac
47401 gctctccaag gaccacctgt gactgcgcac agatactcat cctctcttcc aaatctgatc
47461 ataaaaaact gacaaggcag gtccacgacg aacatagctc cttgtggctt ccacatctcc
47521 ctcaacctcc tgttcaaatt agcaattggg atattccgtc ctaaaacttt caccacctg
47581 cactgcttcc acatactatt catcacatcc aacacttccc taccaatcgt tataaccgat
47641 tcaccatctg ctccattcgg gaactctaca cgcactctcg aagaaacaaa ctcatcatcc
47701 atcaaactct ctgggcaaag cctccctcct gcatctgtac ccactactct cgcgcgccaa
47761 gatcccttcc catccggcgg atctcatgga ggccgtacac cctctccgat ctctgtcata
47821 gtagcatcgt gatcaccctg aacctagacc gtcactgtta cccaactctt ctgttcaact
47881 atctgaatgt agtagtttag catgatttgt ttgtgctggt tagataaatt caacataaga
47941 tgaattaaat tactttgttt cctgtaatt gataaactaa ttcagtgata ttatttgata
48001 aactaatgat ttcttgggtt tggtgataca acattactat ttatgatttt tgtttatcta
48061 atgttataaa gaaagcccca gcccaaaaaa aacgtatcgt cccatttata aaaaaaaat
48121 caactggt
```

//

Revised: July 5, 2002.

[Disclaimer](#) | [Write to the Help Desk](#)
[NCBI](#) | [NLM](#) | [NIH](#)

Jan 21 2003 18:08:12

=> fil reg; d que 17

FILE 'REGISTRY' ENTERED AT 15:18:25 ON 24 JAN 2003

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2003 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file provided by InfoChem.

STRUCTURE FILE UPDATES: 23 JAN 2003 HIGHEST RN 480990-41-8

DICTIONARY FILE UPDATES: 23 JAN 2003 HIGHEST RN 480990-41-8

TSCA INFORMATION NOW CURRENT THROUGH MAY 20, 2002

Please note that search-term pricing does apply when conducting SmartSELECT searches.

Crossover limits have been increased. See HELP CROSSOVER for details.

Experimental and calculated property data are now available. See HELP PROPERTIES for more information. See STNote 27, Searching Properties in the CAS Registry File, for complete details:

<http://www.cas.org/ONLINE/STN/STNOTES/stnotes27.pdf>

L7 1 SEA FILE=REGISTRY ABB=ON UGCAAAGGUAAAAUUAUUUUUUUAUUUGUGUAGCUUAAC
AUGUUUUUGUGUGGAACAAGAUACAKAACAAUGCCAUACCCUCUGCUUAGCUUCUGUAUCCAC
AAGUAUUUGUACUGGUAGU|ACUACCAGUACAAAUACUUGUGGAAUACAGAAGCUAAGCAG
AGGGUAUGGCAUUGUUMUGUAUCUUGUCCACACAAAACAUGUUAAGCUACACAAAUAAAAA
UAAUUUACCUUGCA/SQSN

=> d rn cn sql kwic nte lc 17

L7 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2003 ACS

RN 248237-81-2 REGISTRY

CN GenBank AL132969 (9CI) (CA INDEX NAME)

SQL 99814

SEQ 13601 ttacactcac tgcaaaggta aattattttt tatttgtgta gcttaacatg

13651 ttttgtgtgg aacaagatac agaacaatgc cataccctct gcttagcttc

13701 tgtattccac aagtatttgt tactggtagt caaacaatac tctgatttgt

HITS AT: 13611-13730

NTE doublestranded

LC STN Files: GENBANK

=> fil hom

FILE 'HOME' ENTERED AT 15:18:45 ON 24 JAN 2003

*- Genbank record is 45 pages long (!)
so didn't print it.
URL provided via e-mail.
Chromosome
3*